Journal of Phonetics 71 (2018) 306–325

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Marking 50 Years of Research on Voice Onset Time, eds. Cho, Docherty & Whalen

Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops

Jalal Al-Tamimi *, Ghada Khattab

Speech and Language Sciences, Newcastle University, NE1 7RU, United Kingdom

A R T I C L E I N F O

Article history:
Received 15 November 2017
Received in revised form 11 September 2018
Accepted 27 September 2018

Keywords:
Voicing
Gemination
Fortis
(non-)Temporal
Arabic
Linear Mixed Effects Modelling
Random Forests

A B S T R A C T

This study explores which acoustic correlates best distinguish the voicing contrast in Lebanese Arabic, a language with a two-way voicing contrast that occurs with both singleton and geminate stops. The required timing, phonation and articulatory strength settings for each contrast act synergistically in the voiceless set, but it is unclear how the contrasting requirements for voiced geminates are implemented. Twenty adult speakers were recorded producing target words with medial singleton and geminate stops preceded by long and short vowels. Several temporal and non-temporal measures (duration, VOT, percent voicing, f0, F1, intensity, $H1^*–H2^*$) were taken in the surrounding vowels and in the closure and release phases. Results show that closure duration is the most important cue for distinguishing both voicing and gemination. Active and passive voicing patterns in the closure of voiceless and voiced stops point to [voice] as the main distinctive feature, with [tense] as a secondary feature for voiceless and for geminate stops, with a graded effect. Non-temporal correlates show geminates to have increased tension and creak. Crucially though, voicing is still active in voiced geminates, and release properties have more in common with lenis than fortis languages, leading to a complex profile for this marked category of sounds.

1. Introduction

Voicing and gemination have had a parallel history in terms of the debates around their phonological representation and the gradual untangling of their varied phonetic manifestations, especially with the advance of acoustic and articulatory techniques. One key factor contributing to the difficulty in providing a comprehensive or unified account for each of these terms is that the contrasts embedded in them are based on a complex web of articulatory timing, muscular effort and aerodynamic control; the primacy of each as well as the interplay between them varies within and across languages. In terms of voicing in stops, the timing of vocal fold vibration during a stop's closure phase and release have proven key for distinguishing Voiced and Voiceless[1] stops in many languages, leading Voice Onset Time (VOT) to be adopted in countless studies as the main correlate for the voicing contrast in many languages (see Abramson & Whalen, 2017 for a review). In some languages, however, VOT cannot be used as the sole cue for voicing distinction, especially in cases where the voicing contrast is based on 3- or a 4-way contrast (e.g., Kim, Beddor, & Horrocks, 2002; Lisker & Abramson, 1964; Shimizu, 1996). In terms of gemination in stops, closure duration has been shown to play a major role in distinguishing singleton and geminate stops in many languages (Al-Tamimi & Khattab, 2011, 2015; Arvaniti & Tserdanelis, 2000; Esposito & DiBenedetto, 1999; Ham, 2001; Hassan, 2002; Khattab & Al-Tamimi, 2008, 2014; Podesva, 2000; Ridouane, 2010; Tserdanelis & Arvaniti, 2001). In other

* Corresponding author.
E-mail addresses: Jalal.Al-Tamimi@newcastle.ac.uk (J. Al-Tamimi), Ghada.Khattab@newcastle.ac.uk (G. Khattab).

[1] In this study we use the terms 'Voiced' and 'Voiceless' to refer to phonological voicing and to separate them from actual phonetic voicing, following Docherty (1992). We have refrained from labelling the stops as 'lenis' and 'fortis' instead, partly due to the contrast in Lebanese Arabic being based on true voicing, but more importantly due to the study looking at the geminate contrast as well. Given that the terms 'lenis' and 'fortis' have been used separately in each of the literatures on voicing contrast and gemination, and in each case either to describe the underlying phonological contrast or the phonetic implementation, it was important to keep the two separate.

https://doi.org/10.1016/j.wocn.2018.09.010
0095-4470/

languages, however, there is disagreement over whether the singleton-geminate contrast is based on duration or articulatory strength, especially in initial position where closure duration may be masked; this has led researchers like Louali & Puech (1994) and Ouakrim (2003) to argue for a tense contrast as a phonological feature for Tashlhiyt Berber while Ridouane (2010, for Tashlhiyt Berber) and Abramson (1986, for Pattani Malay) argued for a durational contrast even in initial position.

The primary lexical contrast in each of voicing and gemination is therefore not necessarily based on the surface value of the labels, i.e. actual presence or absence of voicing in the first, and actual lengthening or shortening in the second, but may be driven by other aspects of their production and may vary widely from one prosodic context to the next. This has led to several models of representation of glottal and supraglottal activities involved in their production, and of timing, which plays a key role in both (e.g., Ham, 2001; Jansen, 2004; Keating, 1984; Kohler, 1984; Lisker & Abramson, 1964). Some models emphasise the role of articulatory strength over timing in the representation of each of voicing and gemination (e.g. [±tense] or [fortis]/[lenis]), which creates a conundrum when looking at consonants that are both Voiced and geminate. Voiced stops are typically described as having lax or lenis properties, and often undergo shortening in 'true voice' languages in order to maintain voicing (Jansen, 2004; Jessen, 2001; Kohler, 1984). Geminate stops are typically described as having tense or fortis properties, and have closure durations that are typically between 1.5 to 3 times as long as their singleton counterparts (e.g., Aoyama, 2002; Ham, 2001; Ladefoged & Maddieson, 1996; Ridouane, 2007; Westbury and Keating, 1986). This puts the fortis/lenis properties at odds with each other in Voiced geminates and raises a question regarding both their phonological representation and phonetic implementation.

In this study we look at the benefit of exploring the patterns of stop voicing in Lebanese Arabic (LA), a language with both a voicing and a geminate contrast. The aim is to explore how phonation, timing and articulatory strength interact in the realisation of Voiced geminates, which present a sub-optimal environment for voicing production due to their required length and tense setting. We first provide a brief overview of phonetic and phonological aspects of the voicing contrast across languages, followed by a description of voicing in geminates. We then move on to the main study, which looks at durational and non-durational correlates in the production of each of voicing and gemination contrasts in medial position in LA. In terms of duration, we explore how the timing requirements for the [voice] feature interact with those of the [long] feature required for geminates, and their implication for a VOT measure. In terms of non-durational correlates, we look at how the phonetic implementations of fortis/lenis features in voicing interact with those of tense/lax features in gemination and consider their implications for the implementation of Voiced geminates. In doing so, we explore how conflicting timing requirements are met for each of the lenis feature that is typically associated with Voiced stops and the tense feature that is typically associated with geminates.

1.1. Voicing contrast across languages

Languages which display a two-way phonological voicing contrast typically fall into one of two main groups, *voicing* or *aspirating* languages, depending on how the contrast is implemented in terms of the timing of voicing in relation to supraglottal events (e.g., Abramson & Whalen, 2017; Beckman, Jessen, & Ringen, 2013; Butcher, 2004; Jessen, 2001; Jessen & Ringen, 2002; Lisker & Abramson, 1964; Ohala, 1997). Voiced stops in voicing languages generally show voicing lead, i.e. vocal fold vibration in their closure duration, while Voiceless stops are typically voiceless unaspirated; aspirating languages, on the other hand, typically show similar voiceless patterns in their closure duration for their Voiced stops as is found for Voiceless stops in true voicing languages, while their Voiceless stops are aspirated. The small body of research that exists on voicing patterns in Arabic suggests that the patterns are typical of what is described for true voicing languages (e.g., Al-Ani, 1970; Khattab, 2002; Kulikov, 2016; Yeni-Komshian, Caramazza, & Preston, 1977). The preponderance of languages with patterns that can be divided into one or the other group above led Lisker and Abramson (1964) to establish a classic typology for the voicing contrast in word-initial stops based on VOT as the main acoustic correlate. The three categories (prevoiced, voiceless unaspirated, and voiceless aspirated) were deemed to be sufficient for distinguishing the contrast in the majority of languages. For instance, the majority of Germanic languages contrast voiceless unaspirated with voiceless aspirated stops; on the other hand a large proportion of Romance languages contrast prevoiced with voiceless unaspirated stops (Abramson & Whalen, 2017; Butcher, 2004; Cho & Ladefoged, 1999). This has led researchers to argue for the feature [voice] as representing the laryngeal contrast for both true voice and aspirating languages (e.g., Keating, 1984; Kingston & Diehl, 1994, 1995; Wetzels & Mascaró, 2001).

The distinction is not, however, always clear-cut. For instance, researchers on aspirating languages like English have found prevoicing alongside voiceless unaspiration in Voiced stops (e.g., Davidson, 2016 for American English; Docherty, 1992 for British English; Lisker & Abramson, 1964 for American English). Equally, studies on voicing languages like Arabic have found only partial prevoicing in some Voiced stops and aspiration in Voiceless stops (e.g., Flege & Port, 1981, for Saudi Arabic; Kulikov, 2016, for Qatari Arabic). And while the patterns of voicing and aspiration have mainly been looked at in absolute initial position, these patterns are more complex in sentence- or word-medial position, with some aspirating languages showing evidence of passive voicing for Voiced stops which are meant to be phonetically voiceless (Beckman et al., 2013; Deterding & Nolan, 2007; Jessen, 1998, 2001; Kohler, 1984). Crucially though, these patterns of passive voicing have been found to be qualitatively different from active voicing. For instance, while passive voicing in intervocalic lenis (Voiced) stops in German shows voicing amplitude which gradually decreases, active voicing in Russian intervocalic Voiced stops does not show this amplitude drop (Ringen & Kulikov, 2012). Jansen (2004) further notes that there is a difference between the fortis (Voiceless) stops in voicing languages and lenis (Voiced) stops in aspirating languages, with the former never exhibiting passive voicing. This led some researchers to argue that the [voice] feature alone is not sufficient to describe the voicing contrast in all languages (e.g., Keating, 1984; Kingston & Diehl, 1994, 1995; Kohler, 1984); others suggested [spread glottis] as more suitable for the description of the laryngeal contrast for stops in aspirating languages (Beckman et al., 2013; Harris, 1994; Honeybone, 2005; Iverson & Salmons, 1995; Jansen, 2004; Jessen & Ringen, 2002). This allows a focus not only on the timing of voicing, but also on how languages differ in the tension applied to glottal and supraglottal events during stops production.

Beckman et al. (2013) argue that laryngeal features are privative rather than binary, i.e. that they are defined by the presence or absence of a gesture, rather than both values of a feature having equal status. This allows the separation between a language like Russian where [voice] contrasts with [Ø] (where [Ø] refers to no laryngeal specification), in which [voice] will show active voicing in medial position, with a language like English or German where [spread glottis] contrasts with [Ø] (voiceless aspiration vs. no laryngeal specification), where [Ø] may show passive voicing in medial position. Beckman, et al. (2013) go further to suggest that more than one of these laryngeal features may be needed to describe voicing patterns in languages with seemingly similar laryngeal contrasts, and that different languages can assign different numerical values to these features (e.g. 1–9, whereby 1 is inactive and 9 is highly active) depending on how their stops pattern in different prosodic contexts. They give the example of Icelandic and German, which both contrast [spread glottis] with [Ø], but whereby the lenis (Voiced) stops show passive voicing in German but never in Icelandic. Icelandic lenis stops were found to be produced with slight active glottal spreading, requiring an assignment of a higher numerical feature to their laryngeal feature than their German counterpart ([5sg] [Øvoice] for Icelandic and [1sg] [Øvoice] for German).

Passive devoicing is also typically observed in the closure phase of a phonetically voiced stop due to the increased supraglottal air pressure during the constriction, which affects the transglottic pressure needed to maintain voicing by equalising pressure between the supraglottal and glottal cavities (Jaeger, 1978; Jansen, 2004; Westbury, 1983). Voicing in stops is therefore hard to maintain, which could lead to their devoicing over time or to shortening in order to approximate the closure duration to the time it takes for pressure drop at the glottis. One consequence is that Voiced stops are rarer than Voiceless ones and more typically devoiced across languages (Blevins, 2004; Jaeger, 1978; Kohler, 1984; Westbury, 1983), unless speakers adopt particular strategies to keep the conditions optimal for voicing. This includes lowering the larynx, advancing the tongue root or raising the tongue and the soft palate in order to actively increase the size of their pharyngeal cavity (Halle & Stevens, 1971; Jansen, 2004; Kohler, 1984; Perkell, 1971; Solé, 2009; Stevens, 1998; Westbury, 1983).

The strategies for active voicing above highlight supralaryngeal aspects of the voicing contrast and their contribution to the timing of voicing. For instance, Kohler (1984) sees the voicing contrast as more aptly based on the [±fortis], and stresses the role of co-ordinating the oral, glottal, and velopharyngeal valves in the production of fortis and lenis consonants. Differences between fortis and lenis consonants are evident at each of these valves, with both temporal and non-temporal consequences. In terms of the oral valve, a narrower and tighter constriction is formed faster in the oral cavity for fortis than lenis consonants, with repercussions for preceding vowel duration and quality. A faster movement of the velopharyngeal valve is evident too, with tighter blocking of the nasal cavity. At the glottis, there is tightening, or abduction of the vocal folds compared with slackening and abduction for lenis consonants. The combination of these features leads to fortis (Voiceless) stops being more auditorily salient because they have higher intensity at certain places in the acoustic signal. This energy typically decays at the approach to the closure of the stop and builds up towards the release, and tends to be more abrupt. Jessen (2001, p. 244), and reference therein) presents a similar account based on a [+voice] and [+tense] distinction. According to his auditory model, a [+tense] stop will show a longer closure phase, a longer release phase, a shorter preceding vowel and longer following vowel, increased f0 and F1 frequencies and a higher $H1–H2$ amplitude difference. The latter is due to breathiness resulting from vocal fold stiffness or a larger glottal opening gesture, though creak/laryngealisation can also lead to lower $H1–H2$ (Garellek, 2012; Hanson, Stevens, Kuo, Chen, & Slifka, 2001; Jessen, 2001; Keating, Garellek, & Kreiman, 2015; Klatt & Klatt, 1990; Kuang & Keating, 2012). On the other hand, a [+voice] stop will display a shorter closure phase, longer preceding and following vowel durations, lowered f0 and F1 frequencies at the onset of the following vowel, with lowered $H1–H2$ amplitude (unless the stops are breathy voiced) (Castleman & Diehl, 1996; Kingston & Diehl, 1995; Kingston, Diehl, Kirk, & Castleman, 2008).

In sum, due to the voicing contrast being based on much more than phonation, it is perhaps not surprising then that phonological representations of voicing which are based on the [±voice] feature have been contested, and features such as [fortis], [lenis] and [spread glottis], which better represent the contrast in some languages, have been suggested (Chomsky & Halle, 1968; Kohler, 1984; Wetzels & Mascaró, 2001). Both the binary aspect and the label have been contested where they proved inadequate in describing the phonetic basis of the phonological contrasts embedded in them (Beckman et al., 2013; Harris, 1994; Honeybone, 2005; Iverson & Salmons, 1995; Jessen & Ringen, 2002). In Section 1.2 we draw a parallel between this and the temporal and non-temporal aspects of the geminate contrast.

1.2. Voicing in geminate consonants

Languages with geminates have a contrast between phonologically short and long consonants. Just like early usage of the feature [±voice] in the representation of the voicing contrast, early phonological representations have used [±long] as the distinctive feature between singleton and geminates (Chomsky & Halle, 1968). As was discussed for voicing, however, both the phonology and the phonetic implementation are richer and more varied. In terms of phonetic implementation, consonant duration has proven to be the major cue for gemination in a large proportion of languages (Al-Tamimi & Khattab, 2011, 2015, Arvaniti, 1999, 2001; Arvaniti & Tserdanelis, 2000; Blevins, 2004; Ham, 2001; Hassan, 2002; Khattab & Al-Tamimi, 2008, 2014; Lahiri & Hankamer, 1988; Local & Simpson, 1999; Payne, 2005, 2006; Tserdanelis & Arvaniti, 2001). Temporal properties of the surrounding vowels have been found to act as a secondary correlate to the voicing contrast, with preceding vowels typically shorter before geminate than singleton consonants, but with language-specific differences in the extent and the direction of the change (Al-Tamimi & Khattab, 2011, 2015; Cohn, Ham, & Podesva, 1999; Esposito & DiBenedetto, 1999; Ham, 2001; Hassan, 2002; Homma, 1981; Khattab & Al-Tamimi, 2008, 2014; Lahiri & Hankamer, 1988). In fact, some languages like Japanese have been found to lengthen vowels before geminates (e.g., Kawahara, 2015; Kingston, Kawahara, Chambless, Masha, & Brenner-Alsopa, 2009).

Non-temporal manifestations have also been reported in a variety of studies. These include spectral properties of the preceding and/or following vowels (e.g., increased intensity and lowered f0 and F1 frequencies at the onset of the following vowel, but increased f0 and lower F1 frequencies at the offset of the preceding vowel, as well as decreased $H1$–$H2$ amplitude difference, as seen in our previous results on fricatives, (e.g., Al-Tamimi & Khattab, 2011, 2015; also see, Arvaniti & Tserdanelis, 2000; Esposito and DiBenedetto, 1999; Idemaru & Guion, 2008; Lahiri & Hankamer, 1988; Local & Simpson, 1988, 1999, Payne, 2005, 2006; Ridouane, 2007; Tserdanelis & Arvaniti, 2001). These manifestations suggest a tense/lax distinction that is thought to enhance the perceptual distance between singletons and geminates, but that is also argued to act as a primary correlate in some languages (e.g., Louali & Puech, 1994; Ouakrim, 2003).

When looking at voicing in geminates, some of the tense features described for geminates above may clash with the lenis features required for voicing. The longer duration in these stops leads to increased air pressure behind the place of articulation, often resulting in a “stronger” type of production (Catford, 1977; Jaeger, 1983). The long closure phase is also expected to favour devoicing, and or even complete voicelessness due to equalisation of air pressure in the supra- and sub-glottal cavities, leading to voicing cessation. This is referred to as the Aerodynamic Voicing Constraint (AVC) (Ohala, 1997, pp. 93–94). According to this constraint, and in order to maintain voicing, a speaker will need to voluntarily use an active voicing strategy by employing some of the alternative strategies described in Section 1.1 (e.g., Jansen, 2004; Kohler, 1984; Solé, 2009; Westbury, 1983). Similarly, vowels preceding Voiced stops are typically longer and have lower f0 offset and intensity, so this may clash with the typical shortening and f0 raising seen in a geminate environment; these effects may also apply in following vowels especially in iambic contexts.

It is therefore not surprising that Voiced geminates are considered marked (Blevins, 2004; Hayes and Steriade, 2004; Ohala, 1983; Westbury & Keating, 1986), and that they are rarer compared with their Voiceless counterparts. One of the few studies to have looked at voicing in languages with contrastive voicing and gemination is by Butcher (2004), who explored voicing in intervocalic bilabial stops in Italian (amongst other languages). Acoustic measures of the closure duration were taken and intraoral pressure was measured using a 2 mm plastic catheter inserted between the lips of the speakers. One interesting finding was that the closure duration for fortis (Voiceless) stops was significantly different from that of the lenis (Voiced) for only the singleton pair, perhaps showing the primacy of closure lengthening in the geminate pair. In terms of pressure differential, fortis stops had higher intraoral pressure than lenis stops for both singleton and geminate pairs, with a more pronounced difference for the geminate pair. The singleton fortis-lenis pair therefore differed in duration and peak pressure, while the geminate fortis-lenis pair differed mainly in pressure.

Jessen (2001, p. 270) described gemination as a phonologisation of the [+tense] feature through the lengthened closure phase. But under this proposition, it is still not clear how Voiced geminates survive as [lenis] and [tense]. As Butcher’s study above showed, they can have lower intraoral pressure than their fortis counterparts, but a similar duration. If the longer closure phase prohibits maintaining voicing for a long time, a stronger degree of devoicing is expected. If the proportion of voicing in the closure phase, on the other hand is not different from that of the singleton, then speakers must be actively using some strategies to allow them to maintain voicing, such as shortening the closure phase, marking voicing at the boundaries of the stop, etc. (e.g., Castleman & Diehl, 1996). We still do not fully understand the temporal and non-temporal properties that make up these marked sounds, or how they pattern cross-linguistically.

1.3. The aim of this study

This paper reports on the acoustic correlates of the voicing contrast in singleton and geminate stops in LA in medial position. A focus on a voicing language allows us to investigate how speakers maintain voicing, especially in phonologically long consonants. A focus on medial position allows us to look at active and passive voicing patterns and refine our analysis of the laryngeal specifications that best describe the LA voicing contrast. A focus on gemination allows us to explore how voicing patterns interact with the phonetic implementations of phonological length, which not only include durational contrast, but also differences in phonation and articulatory force. In this respect, this is the first study on Arabic to look at the phonetic implementation of four categories crossing voicing and gemination and to combine temporal and non-temporal acoustic correlates of the voicing contrast. Specifically, we aim to evaluate the degree of (dis)similarity between geminates and fortis/tense consonants in terms of the effects of the longer closure phase on the degree of (de)voicing and the consequences for the other acoustic correlates investigated here. If geminates are truly fortis/tense, we would observe the same effects on the acoustic correlates, mainly longer closure and release phases, shorter V1 and longer V2 durations, more devoicing in the closure phase (potentially affecting VOT patterns for the Voiced category), higher f0 and higher F1 at the offset of V1 or onset of V2 (depending on stress patterns), and increased $H1$-$H2$ amplitude difference. However, previous results for geminates (e.g., Al-Tamimi & Khattab, 2011, 2015; Arvaniti & Tserdanelis, 2000; Idemaru & Guion, 2008; Ridouane, 2007) point to opposing results for f0, F1 and $H1$–$H2$ effects, especially in V2. Moreover, many of these patterns clash with what is expected for Voiced or lenis consonants, making it hard to predict patterns for Voiced geminate. If geminates in LA do not exhibit the same patterns as fortis/tense stops, differences would be observed in relation to the relevant acoustic correlates which could potentially enable us to separate aspects of the fortis-lenis distinction that are mainly due to laryngeal activity from ones that are manifestations of supra-laryngeal events.

2. Experimental/Materials and methods

2.1. Speakers and corpus

Twenty Lebanese speakers (ten males) aged between 18 and 40 and with no reported history of speech and/or language disorders were recruited from Beirut. They all spoke the Greater Beirut dialect and were university-educated. Due to the multilingual nature of Lebanon, they were all exposed to English and French and knew Modern Standard Arabic through education. Participants were instructed to produce a list of real words (n = 410) that fit into one of the following syllable structures: ˈCVCVC (e.g. /ˈʕadad/ "number"), ˈCVːCVC (e.g. /ˈʕaːded/ "counting"), ˈCVCːVC (e.g. /ˈʕadːad/ "he enumerated"), ˈCVːCːVC (e.g. /ˈʕaːdːe(h)/ "having counted") and CVˈCːVːC (e.g. /ʕaˈdːeːd/ "counter"). We aimed to obtain a representative type of contrasts available in Arabic. The first four syllable structures are trochaic and allow for a four-way durational contrast available in (LA), with both short and long vowels preceding the singleton and geminate consonants; the last one is iambic and shows a pattern of a stressed geminate with a following long vowel. All consonants in LA in both singleton and geminate contexts (C/Cː hereafter) were produced, but here we focus only on stops (98 words per speaker): /b, t, tˤ, d, dˤ, k/.

Near minimal-pair sets were used with the medial C/Cː being one of the stops above, with preceding and following target vowels being /a/ or /aː/ (with /aː/ being frequently realised as [eː] or [ɛː] due to Imāla; Nasr, 1960). Given the length of the full corpus (wordlist alongside spontaneous speech, which is not presented here), we did not elicit repetitions for each word but there were two or three examples for each consonant in a given context. Similarly, we did not use a carrier sentence, but with the target sounds in word medial position this did not affect the measurements of any of the stops phases. Recording sessions lasted around one hour per speaker. Randomisation and fillers (*n* = 40) were included and we used the Modern Standard Arabic script without vowelisation and presented the words in their dialectal form. The participants were asked to produce the words in their own variety and in an informal style at a steady rate. The speech material was recorded in a quiet room, using an R9 solid-state recorder with a SONY MS957 Uni-directional Stereo Electret Condenser microphone (frequency response 50–18000 Hz), and digitised at 44.1 kHz, in mono channel and 16-bit quantisation.

2.2. Semi-automatic segmentation of the data

The corpus was semi-automatically segmented into C(onsonant) and V(owel) intervals using the package STK (Pellegrino & André-Obrecht, 2000). These intervals with their timestamps were transferred into TextGrid and used in Praat (Boersma & Weenink, 2009) for additional data processing. A total of 1960 words were elicited. 171 tokens were discarded due to noise or technical error, leaving a total of 1793 words for subsequent analyses. Auditory transcription of all words and manual acoustic correction of the boundaries created by the semi-automatic system were done by the first author, with the data being rechecked by first and second authors for accuracy. The following criteria were used in positioning the boundaries between segments/portions (Fig. 1):

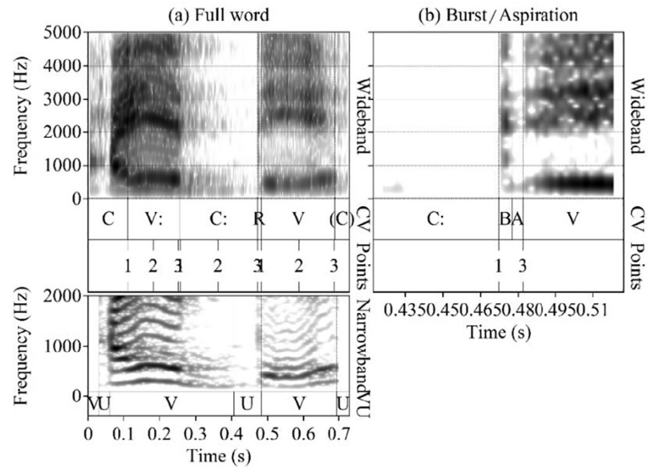

Fig. 1. Segmentation of the word /ʕaːdːe(h)/ "having counted" produced by a female speaker, with the full word (in a) and the Burst/Aspiration portion (in b). The following are seen in both: wideband spectrogram, and segmentation into C, V, R, B and A (consonant, vowel, release, burst and aspiration, respectively, tier CV), points of measurements (tier points). In (a), a narrowband spectrogram and boundaries for the voiced and unvoiced frames (V and U, respectively) are shown. See text for more details.

1. Vowel Onset/Offset was determined using any rise/fall in amplitude from previous/following consonant and the appearance of a homogenous F1–F4 formant structure. We did not use the first glottal cycle's appearance in order to ensure consistency across manners of articulation (Fig. 1a, tier CV, symbols V: and V). In the case of words ending with a vowel, a separate boundary was drawn to signal any voiceless portion after the end of the second vowel (Fig. 1a, tier CV, and symbol (C)) as these portions do not seem to contribute to the perception of vowel duration (see Nakai, Kunnari, Turk, Suomi, & Ylitalo, 2009).
2. The medial stop was segmented into various portions (Fig. 1a and b)
   a. Closure phase (Fig. 1a, tier CV, symbol C): the onset was determined as the end of the preceding vowel. Any portion of weak formant or voicing lead into the closure phase were included in the medial stop. The offset of the closure phase was decided as the onset of a visible burst indicating the release of the stop.
   b. Release phase (Fig. 1a, tier CV, symbol R): the onset was determined as the first visible burst and the offset as the onset of the following vowel[2]. The Release phase was taken as our positive VOT measure (see Section 2.3.3). This portion was then segmented into two intervals for a more fine-grained analysis of the release properties as a function of place, voicing and gemination:
      i. Burst (Fig. 1b, tier CV, symbol B): its onset was determined as the first visible burst and its end as the last visible burst that is separated by less than 5 ms from its previous burst.
      ii. Aspiration (Fig. 1b, tier CV, symbol A): its onset was determined as the last visible burst and its offset as the beginning of the onset of the following vowel. This included any parts of frication after the "burst".

2.3. Acoustic analyses

Multiple acoustic correlates were automatically obtained using a Praat script designed by the first author and adapted from Al-Tamimi (2017) and Al-Tamimi & Khattab (2015). Below

[2] As this was taken as our positive VOT measure in this study, it is important to point out that it was possible for release phases in this study to contain voicing (see results in Fig. 4), as the vowel rather than the onset of voicing were taken as the end of the release.

Table 1
Summary of acoustic correlates per category of sounds.

Medial consonant	Closure phase	Release phase	Preceding vowel	Following vowel
VOT	Duration %Voicing	Duration Duration (burst) Duration (Asp) %Voicing Intensity Onset Intensity Offset	Duration Intensity Offset F0 Offset F1 Offset $H1^*–H2^*$ Offset	Duration Intensity Onset F0 Onset F1 Onset $H1^*–H2^*$ Onset

we describe the acoustic correlates chosen in this study for their demonstrated role in both voicing and gemination contrasts. A total of 19 acoustic temporal and non-temporal correlates were used (see Table 1 for a summary). Before performing any of the analyses summarised below (see Sections 2.3.2–2.3.7), and to reduce errors obtained by the automatic analyses, an accurate estimation of the frame position at the onset/midpoint/offset was done (see Fig. 1a and b, tier Points, 1, 2 and 3 respectively) following the procedure developed in Al-Tamimi (2004, 2007).[3]

2.3.1. Adjustment of measurement points

The duration of a complete glottal cycle was estimated based on f0 computation (Section 2.3.5). F0 frames were reanalysed by a cross-correlation PointProcess. This glottal cycle ranged over 8–10 ms for males and 4–6 ms for females. The original onset/mid/offset positions were adjusted by up to 2–3 ms, to match the time of maximum intensity occurring within the length of a complete glottal cycle (see Fig. 1a and b, tier Points, positions 1, 2 and 3 for onset, midpoint and offset, respectively). This was done in order to reduce errors in automatic extraction. Durational and voicing measures were reported using the original positions, with all remaining ones using the adjusted positions.

2.3.2. Durations (ms)

The duration was obtained from the original onset to the original offset (Fig. 1a and b, tier CV). These were obtained for the closure phase of the medial stop, the preceding and following vowels, the full release phase (Fig. 1a, tier CV, symbol R), and the burst and aspiration phases (Fig. 1b, tier CV, symbols B and A).

2.3.3. Voicing patterns (ms and %)

We developed an automatic procedure[4] to quantify the degree of voicing/devoicing in medial stops (both closure and release phases) that relied on Praat's voicing detection algorithm. Each sound file was first low-pass filtered at 500 Hz in order to remove any influence of weak formants and f0 estimation was carried out (see Section 2.3.5) with the exception of using cross-correlation. F0 contours estimated using the second pass (see below), were reanalysed with a cross-correlation PointProcess to estimate the length of a complete glottal cycle. We then used Praat's VUV (VoicedUnVoiced) function with an average duration of a complete glottal cycle adapted to each speaker, while the minimum of 20 ms for continuous voiced or unvoiced interval was kept to its default. This procedure created a new TextGrid with the boundaries around Voiced and Unvoiced frames (V and U respectively). We then carried out a manual check of this automatic estimation by low-pass filtering the sound to 2000 Hz and looking at both a narrowband spectrogram[5] and f0 tracks (Fig. 1a, tier Narrowband and VU). In the example in Fig. 1, there is low frequency activity at the end of the closure phase, but it is too low in amplitude to qualify for a "true" voiced portion.

We then computed two measures of voicing: The (positive and negative) VOT (following Abramson & Whalen, 2017) and the proportion of voicing in both closure and release phases separately with respect to their duration. In the example shown in Fig. 1a, and following the 50% voicing threshold, the VOT was −216 ms, with the proportion of voicing in the medial geminate /d/ being 69% in the closure phase and 0% in the release phase. While VOT was calculated using the 50% voicing criteria as suggested in Abramson and Whalen (2017), we also looked at the raw proportion of voicing in each of the closure and release in order to explore whether it offered a more fine-grained distinction between Voiced singleton and geminate stops.

2.3.4. Intensity (dB)

The default setting in Praat (32 ms Gaussian window, 100 Hz pitch floor, 5 ms time step, and cubic interpolation) were used to obtain the intensity at the offset of the preceding vowel, the onset and offset of the release phase and the onset of the following vowel.

2.3.5. Fundamental frequency (Hz)

The fundamental frequency f0 was estimated following the procedure described in Al-Tamimi, (2017) and Al-Tamimi & Khattab (2015) and used the two-pass method (Hirst, 2011), with the aim to reduce errors in f0[6] computation. The first pass relies on Praat default settings: 5 ms time step, 40-ms Kaiser2 window, floor and ceiling = 75–500 Hz respectively, and autocorrelation. The pitch contours were obtained for each speaker based on this method. Then the first and third quartiles at 25% and 75% respectively were obtained and were then multiplied by a coefficient: 0.75 and 1.5 respectively. The resultant values were then used as the new floor and ceiling in the second pass. With this second pass, the actual f0 computation was done with a 5 ms time step, and a 30 ms effective Gaussian window length. The floor and ceiling values ranged between 75–100 and 120–250 for males, and between 125–160 and 190–280 for females.

[3] The script is available from https://github.com/JalalAl-Tamimi/Praat-Measurement-points.
[4] The script is available from https://github.com/JalalAl-Tamimi/Praat-Voicing-detection.
[5] The narrowband spectrogram was used to assess whether any low frequency components visible around the first harmonic are true voiced or unvoiced frames. The amplitude of the portion was also looked at during the decision-making. When weak formants were present in the closure phase without low frequency activity, these were not considered as true voicing.
[6] The script is available from https://github.com/JalalAl-Tamimi/Praat-f0-Accurate-Estimation.

2.3.6. F1 (Hz)

Formant frequencies of the vowels preceding and following the medial stops were estimated using Praat's default Burg algorithm, with a 25 ms Gaussian window, a 5 ms time step and interpolation. The sound file was automatically downsampled to 10 kHz for males (or 11 kHz for females) through the Burg algorithm as implemented in Praat. A maximum of five formants were requested in the analyses, and a maximum frequency of 5 kHz for male and 5.5 kHz for female speakers. Then we used Praat's track function to correct errors in the automated procedure. Formant frequencies were then checked, and manual correction carried out to prevent errors in automatic extraction (errors constituted less than 5% of the data). Then the frequencies of the first formant (F1) were obtained at the offset of the preceding vowel and the onset of the following vowel.[7]

2.3.7. $H1^*$–$H2^*$ (dB)

Finally, we computed $H1^*$–$H2^*$ as an acoustic correlate of voice quality. The sound files were first low-pass filtered with an anti-aliasing filter which had a cut-off frequency of 5 kHz for male and 5.5 kHz for female speakers, down-sampled to 10 kHz for male and 11 kHz for female speakers, and pre-emphasized by a factor of 0.98. Intervals 40 ms long were defined to estimate spectra of the vowels. For each sound file, one interval was right aligned at the offset of the preceding vowel and another left-aligned at the onset of the following vowel. Then this 40 ms interval was windowed with a Kaiser2 window function. Then from each interval, a 256-point zero-padded DFT spectrum was computed along with the logarithmic power spectral density, with a bin size of 19 Hz. Following Al-Tamimi (2017) and Al-Tamimi & Khattab (2015), the amplitudes of the first and second harmonics (and of the first and second formants for corrections, see below) were estimated by automatically detecting the highest peaks of a particular harmonic. For $H1$ and $H2$, the maximum amplitude was obtained from f0*0.9 to f0*1.1 and from 2*f0*0.95 to 2*f0*1.05 respectively. Following the recommendations of Iseli, Shue, & Alwan (2007), we estimated the bandwidths based on Hawks & Miller (1995)'s formula. Then, the amplitudes of the harmonics closest to the first two formants were obtained in the region from F1 – 0.5*Bandwidth1 to F1 + 0.5*Bandwidth1 for $A1$ and F2 – 0.5*Bandwidth2 to F2 + 0.5*Bandwidth2 for $A2$. The automatic detection was manually checked to prevent errors (through visual inspection of the spectra and automatically obtaining the amplitudes with Praat's built-in functions). Then the normalisation procedure as developed by Iseli et al. (2007) was implemented in our Praat script to obtain the corrected amplitudes of these harmonics. Both $H1$ and $H2$ were corrected for the boosting effects of both F1 and F2. Amplitude differences were obtained by subtracting $H2^*$ from $H1^*$ ("$H1^*$–$H2^*$") at the offset of the preceding vowel and onset of the following vowel.

[7] It is to be noted that F1 frequencies obtained using the Burg method can be problematic given that the formants detected may be attracted to a strong harmonic (see Shadle, Nam, & Whalen, 2016 for more details). However, the actual testing of the procedure in the arburgh function that the authors used (with 30 ms and 14 poles LPC) may be more equivalent to Praat's "Sound: To LPC (Burg…)" with 14 poles, which we have not used, following recommendation from Praat's manual.

2.4. Statistical analyses

The acoustic correlates summarised above yielded a total of 34,067 measurement points (medial consonant VOT = 1793, closure phase = 3586, release phase = 10758, preceding vowel = 8965 and following vowel = 8965). We adopted a predictive modelling approach (Baguley, 2012; Hastie, Tibshirani, & Friedman, 2009; James, Witten, & Hastie, 2013; Kuhn & Johnson, 2013) to evaluate the degree to which the combination of these correlates can provide a meaningful insight into the voicing contrast in the singleton and geminate stops in LA. To achieve that, the data were first statistically analysed via a confirmatory Linear Mixed Effects Modelling followed by Random Forests as a classification technique. All statistical analyses were performed using the statistical software `R version 3.5.0` (Microsoft R Open, R Core Team, 2018).

2.4.1. Linear Mixed Effects Modelling

Linear Mixed Effects Modelling (LMM) were applied using the package `lme4` (version 1.1-14, Bates, Mächler, Bolker, & Walker, 2015) with each acoustic correlate as the outcome and the following factors as predictors: voicing (Voiceless, Voiced); consonant type (singleton, geminate); sex (male, female); length of V1 (short, long); syllable type (trochaic, iambic); and place of articulation (bilabial, alveolar, pharyngealised, velar). A crossed-random effects structure for speakers and items was used (Baayen, Davidson, & Bates, 2008). Following a maximal specification model (Barr, Levy, Scheepers, & Tily, 2013), by-speaker random slopes were used for: voicing, consonant type, V1 length, syllable type, and place of articulation; these improved the model fit compared to a model without random slopes. We used contrast coding on all fixed effects by centring them to values between -0.5 and 0.5 to allow for a meaningful interpretation of the coefficients (Schielzeth, 2010; Schielzeth & Forstmeier, 2009). Multiway interactions did not improve the model fit for all models ($p > 0.05$), thus main effects models were used here: the results obtained are generalisable over and above all interactions due to contrast coding. Given the way in which LMMs work, it was not necessary to normalise the data between speakers for f0 or F1 frequencies to allow for the predictor "sex" to be modelled appropriately. Below, we will present the graphical results based on the predicted (or fitted) values for each outcome that are adjusted by the statistical model (using the `predict` function in `lme4`). All figures in the following section (Figs. 2–8) were generated using the predicted values from each LMM; these were created using the package `ggplot2` (version 2.2.1, Wickham, 2009) and the package `gridExtra` (version 2.3, Auguie, 2016). We then report the means and standard deviations (SDs) in a four-way contrast (Voiceless singleton, Voiced singleton, Voiceless geminate, Voiced geminate) and provide a pairwise comparison on the predicted (or fitted) values using `pairwise-t-test` in `R`. The pairwise comparisons were corrected for multiple comparisons using the False Discovery Rate (FDR) alpha correction.

2.4.2. Random Forests via conditional Inference trees

After exploring the relationship of predictors on separate acoustic correlates, we used Random Forests as a predictive model that allows the *combination* of predictors to work together, either as main effects or in interaction, to explain the patterns in the data (Hastie et al., 2009; James et al., 2013; Kuhn & Johnson, 2013). Random Forests are used in sociolinguistic research (Tagliamonte & Baayen, 2012), in acoustic cue weighting in perception (Baumann & Winter, 2018; Brown, Winter, Idemaru, & Grawunder, 2014), and on multiplicity of correlates in pharyngealisation (Al-Tamimi, 2017).

We used Random Forests grown via Conditional Inference Trees as they guard against biases introduced by the original implementation that favoured predictors with multiple cut-points and categories, and overestimated variable importance for correlated data (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008; Strobl, Boulesteix, Zeileis, & Hothorn, 2007). Subsampling without replacement provides an unbiased selection process (Strobl, Malley, & Tutz, 2009), and conditional permutations of predictors allow for a controlled evaluation of the importance of variables in classification and/or regression after taking into account their correlations and interactions (Strobl et al., 2009; Tagliamonte & Baayen, 2012).

We used the function `cforest` from the package `party` package (version 1.2–3, Hothorn, Hornik, & Zeileis, 2006; Strobl, et al., 2008; Strobl, et al., 2007) to grow the forests on the combined predictors, with the recommended `cforest_unbiased` control and `mtry` adapted to each forest. The predictors were all z-scored to a mean of 0 and a standard deviation of 1 to normalise for the magnitude of the differences in their scales and to put all predictors on the same level. Following the procedure proposed by Oshiro, Perez and Baranauskas (2012), and implemented in our previous work (Al-Tamimi, 2017), we tuned the number of trees needed to grow a forest with the highest predictive accuracy by growing 15 random forests with `ntree` from 100 to 1500 trees in 100 trees increment[8] and evaluating the predictive accuracy via an AUC (for Area Under the Curve) based comparison using the package `pROC` (version 1.10-0, Robin, et al., 2011). A non-parametric Z test of significance on correlated ROC curves was carried out using the function `roc.test` (following DeLong, DeLong, & Clarke-Pearson, 1988). In the end 500 trees for model A, 500 trees for model B and 600 trees for model C (see below) were enough to grow forests that had the highest predictive accuracy.

To grow the forests, we divided the dataset into a training set (66.6% of the data) and a testing set (33.3% of the data). We then trained three Random Forests. The outcome had a four-way level: Voiceless singleton vs Voiced singleton vs Voiceless geminate vs Voiced geminate and either the full 19 predictors (Model A), 18 predictors without the closure phase (Model B) or 17 predictors without the closure phase or the VOT (Model C). The aim of these three models were to assess which predictor(s) are most predictive of the four-way contrast and what each of the closure phase, VOT and additional correlates are contributing to the contrast (see Section 1.2). Once the three forests were trained, we used the function `predict` (and the specifying `OOB = TRUE` for cross-validation.) to predict the outcomes using the testing set. We then calculated the percent correct classification based on a confusion matrix of the classification results. Following this, we estimated the importance score of each predictor (using `varimp`) and used conditional permutation tests (by specifying `conditional = TRUE`). These conditional permutations create a comparison grid that includes all predictors and performs a conditional comparison, comparable to that in a regression analysis. The variable importance results are representative of the true contribution of each predictor after taking into account correlations and interactions. The figures were generated with `ggplot2` (version 2.2.1, Wickham, 2009) and the `gridExtra` (version 2.3, Auguie, 2016).

3. Results

We first look at the LMM results for each of the acoustic correlates (Section 3.1) then the classification results obtained via Random Forests (Section 3.2). The full results of both sections are also available in the supplementary material available here: https://jalalal-tamimi.github.io/R-Voicing-Gemination-VOT/.

3.1. LMM results

In the following sections, we evaluate the effect of voicing, i.e., Voiced vs Voiceless in the singleton vs geminate contexts. All models improved the model fit except for $H1^{*}–H2^{*}$ at the offset of the preceding vowel (for model comparison results, see supplementary material online). The full statistical results of the optimal models are presented in Appendix A.

3.1.1. Durations (ms)

Duration (ms) results are divided into two parts: 1) preceding vowel (V1), the closure phase (C2) and the following vowel (V2) and 2) the release phase (Rel), as well as the burst phase (Burst) and the aspiration phase (Asp) within it. Results for the former group are presented in Fig. 2, whereas those for the latter are in Fig. 3. As a reminder here and throughout, when the results of the preceding or following vowel are presented, they refer to both short and long preceding vowel (due to differences in syllable structures). This was done to generalise the results, although the statistical results presented in Appendix A are already adjusted for the effects of vowel length and syllable type differences.

Starting with the first group, voicing state affected vowel and consonant durations in opposing manners (see Fig. 2, Table 2 and Appendix A). Within V1, there is an overall statistically significant increase by an average of 24 ms ($p < 0.0001$) in the Voiced compared to the Voiceless context in both singleton and geminate environments. Within geminates vs singletons, the former displayed a statistically significant decrease in V1 duration by an average of 32 ms ($p < 0.0001$). Moving on to the CD, the pattern is reversed, with a statistically significant decrease by an average of 18 ms in CD ($p < 0.0001$) in the Voiced compared to the Voiceless stops. Within geminates vs singletons, the former showed a statistically significant increase in CD by an average of 98 ms ($p < 0.0001$). Finally, with respect to V2 duration, and within Voiced vs Voiceless contexts, there was only a tendency for an increase in V2 duration in the Singleton Voiced compared with the Voiceless context ($p = 0.055$). However, within geminates vs singletons, the

[8] An example script is available from https://github.com/JalalAl-Tamimi/R-Estimating-Number-Of-Trees-RF.

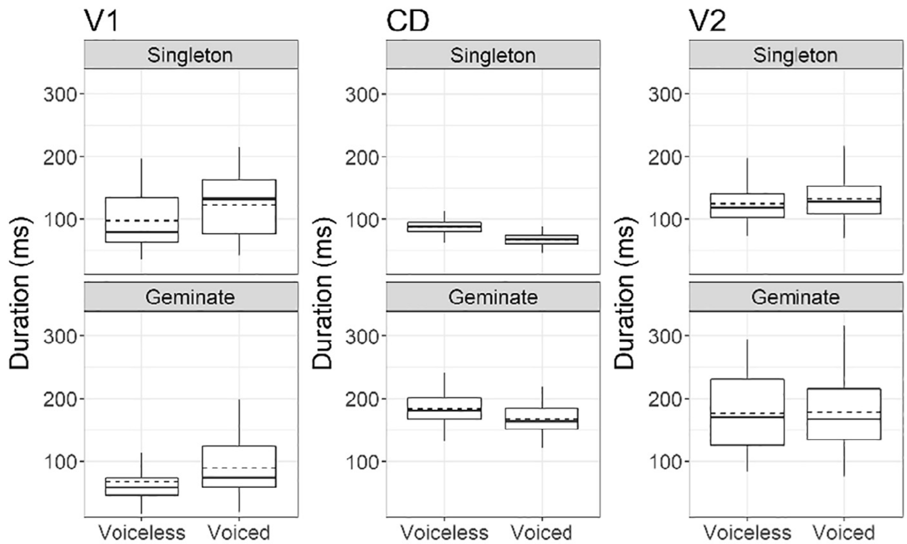

Fig. 2. Duration (in ms) boxplots adjusted by our statistical model for the preceding vowel (V1), the closure phase of the medial consonant (CD) and the following vowel (V2) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

Table 2
Descriptive statistics for the duration (ms) with mean and standard deviation (SD) for the four-way contrast for the preceding vowel (V1), the closure phase (CD), the following vowel (V2), the release phase (Rel), the Burst phase (B) and the aspiration phase (Asp).

	Voiceless Singleton		Voiced Singleton		Voiceless Geminate		Voiced Geminate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V1	97.91	43.42	123.25	49.01	67.10	32.53	89.60	42.91
CD	87.57	10.75	67.97	8.98	184.13	22.59	167.57	20.84
V2	125.17	29.57	131.70	30.93	176.08	55.19	178.42	52.78
Rel	28.62	9.44	11.90	4.84	25.06	8.97	13.96	4.59
B	8.83	2.67	5.03	1.12	9.79	2.69	6.12	1.16
Asp	19.97	7.91	7.89	3.74	15.40	7.35	7.96	3.67

former displayed a statistically significant increase by an average of 49 ms in V2 ($p < 0.0001$).

The duration results presented so far show the combined effect of voicing and gemination, with V1 showing the increased duration following Voiced stops and the shortening effect following geminates. Voiced geminates show that combined effect too, with shorter closures than singleton counterparts but voicing effects remaining clear in V1 duration.

Moving on to the second group, Fig. 3 (and Table 2) shows voicing as having a major impact on the duration of the

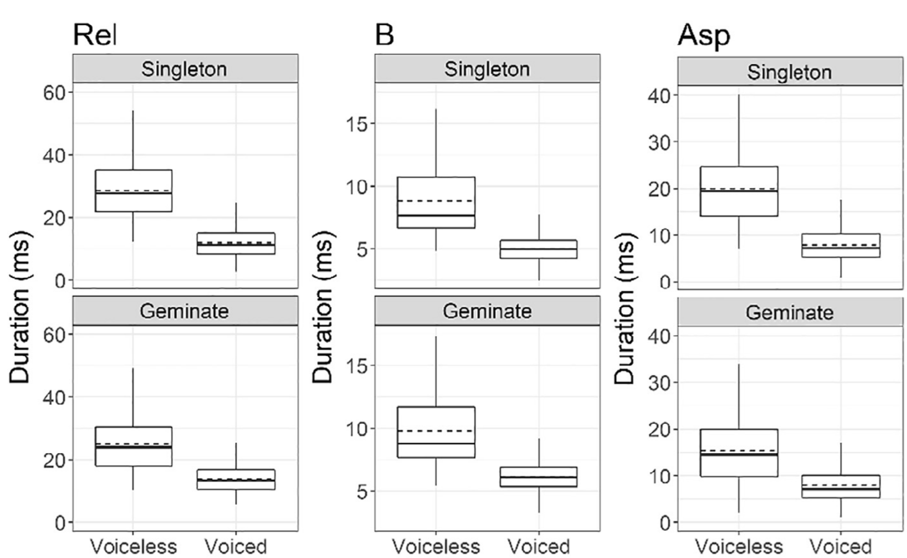

Fig. 3. Duration (ms) boxplots adjusted by our statistical model for the release phase (Rel), the burst phase (Burst) and the aspiration phase (Asp) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

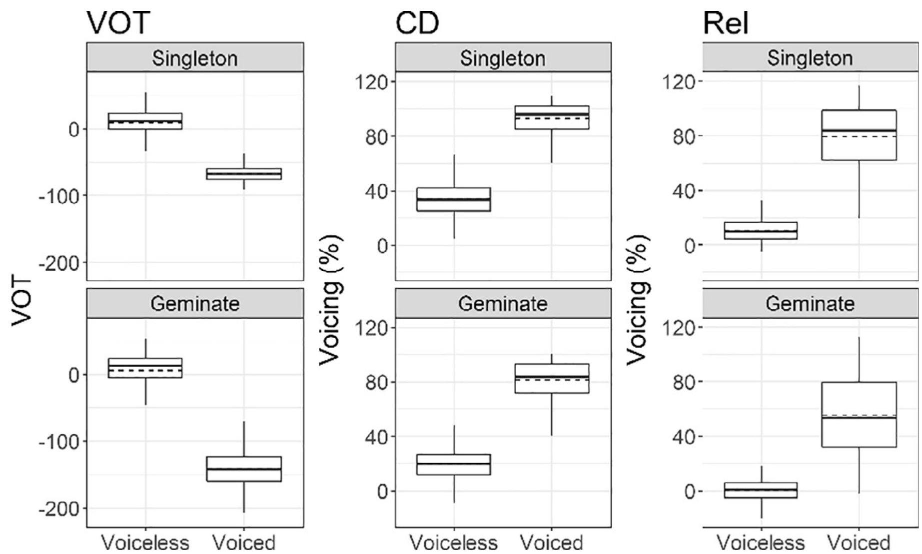

Fig. 4. VOT (ms) and proportion of voicing (%) boxplots adjusted by our statistical model for the closure phase of the medial consonant (CD) and the release phase (Rel) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

Table 3
Descriptive statistics for the voicing patterns with mean and standard deviation (SD) for the four-way contrast for the positive and negative VOT (ms), % voicing in the closure phase (CD) and in the release phase (Rel).

	Voiceless Singleton		Voiced Singleton		Voiceless Geminate		Voiced Geminate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
VOT	8.70	23.26	−67.04	11.91	5.31	29.69	−140.65	28.61
CD	34.35%	13.42%	92.64%	12.34%	19.90%	12.17%	81.04%	13.84%
Rel	10.84%	8.56%	79.56%	23.00%	0.42%	7.87%	55.34%	26.68%

Release, the burst and the aspiration phases (see Appendix A for statistical results). Starting with the release phase (Rel), there is a statistically significant decrease by an average of 14 ms in the duration of Rel ($p < 0.0001$) in the Voiced context. Within geminates vs singletons, there was an opposing effect: in the Voiced context, geminates showed a statistically significant increase by 2 ms in the duration of Rel ($p < 0.0001$) and within the Voiceless, they showed a statistically significant decrease by 3.5 ms in the duration of Rel ($p < 0.0001$). Moving on to the burst phase (Burst), there is an overall statistically significant decrease by an average of 3 ms in the duration of the Burst ($p < 0.0001$) in the Voiced context. Within geminates vs singletons, there was a statistically significant difference recorded, with the geminate showing an increase by an average of 1 ms ($p < 0.0001$). With respect to the aspiration phase (Asp), there is an overall statistically significant decrease by an average of 10 ms in the duration of Asp ($p < 0.0001$) in the Voiced context. And finally, within geminates vs singletons, there was only a statistically significant decrease in the aspiration portion in the Voiceless geminate compared to the Voiceless singleton by 4.5 ms ($p < 0.0001$). It should be noted that all of the duration results, except those for the Burst and Release phase differences within singleton vs geminates, are close to or beyond those observed in the Just Noticeable Difference (JND) in temporal discrimination (Stevens, 1998, pp. 228–229).

The results for the release phase demonstrate voicing effects over gemination, including for the Voiced geminate. Voiceless stops exhibit the short aspiration that is typical of fortis stops in voicing languages while Voiced ones have a very short release before the start of the following vowel. A marginal geminate effect is seen in terms of the increase in number of bursts compared with singletons,[9] but crucially, release duration does not distinguish singleton from geminate stops.

3.1.2. VOT (ms) and voicing patterns (%)

Graphical results are presented in Fig. 4 and Table 3 (see Appendix A for full statistical results) for VOT and the proportion of voicing in the closure (CD) and the release (Rel).

Starting with VOT, the results show a clear distinction between the Voiced and Voiceless stops, with the former having a negative VOT and the latter a positive VOT; Voiced stops had a statistically significant longer negative VOT by an average −110 ms (ranging between −75 ms in the singleton to −146 ms in the geminate, $p < 0.0001$). Within geminates vs singletons, geminates in the Voiced context showed a longer negative VOT than the singleton Voiced by −74 ms ($p < 0.0001$); however, there was a tendency for Voiceless geminates to show a marginal decrease in positive VOT by 3 ms ($p = 0.064$). It is interesting to see that the VOT is not able to clearly distinguish between the singleton and geminate Voiceless categories.

Moving on to the proportion of voicing in the closure (CD), there is a statistically significant increase by an average of 60% in the proportion of voicing ($p < 0.0001$) in the Voiced con-

[9] Overall, the number of bursts in a geminate context were on average 2.5 (SD = 1.2), whereas these were on average 1.5 (SD = 1.1) in the singleton context. Place of articulation had a positive effect of number of bursts, with an average number of 3 bursts (SD = 1.4) in velars; in geminate velars, there were on average 4 bursts (SD = 0.5).

text compared to the Voiceless context, which is comparable to the results of the VOT reported above. Within geminates vs singletons, there was a statistically significant decrease in the proportion of voicing in the geminate context by an average of 13% ($p < 0.0001$) in both Voiced and Voiceless contexts. Even within the Voiceless, geminates showed a decrease in the proportion of voicing lead in the closure compared to singletons by 14% ($p < 0.0001$); a result not captured by the traditional VOT. With respect to the Release (Rel), there is a statistically significant increase by an average of 62% in the proportion of voicing ($p < 0.0001$) in the Voiced context compared to the Voiceless context. Within geminates vs singletons, there was a statistically significant decrease in the proportion of voicing in the geminate context by an average of 17% ($p < 0.0001$).

Overall, the proportion of voicing in the closure phase is a clear acoustic correlate that can be used to distinguish both voicing and gemination contrasts in LA. The patterns found suggest that LA behaves as a true voicing language, as the VOT results show a clear pattern between the Voiced and the Voiceless set. However, the VOT alone was not able to reflect the decrease in the proportion of voicing in Voiced geminates compared with Voiceless ones, nor was it able to differentiate the Voiceless geminate from the Voiceless singleton. The results for the release phase (Rel) support the CD patterns.

3.1.3. Intensity (dB)

The results of the intensity (dB) obtained at the offset of V1, the onset and offset of the Release (Rel) and at the onset of V2 are presented in Table 4 and Fig. 5 (also see Appendix A).

Starting with V1 Offset, there is an overall statistically significant increase in the intensity of V1 Offset in the Voiced context by an average of 3 dB ($p < 0.0001$). Within geminates vs singletons, there was a statistically significant decrease in the intensity at the Offset of V1 by 1.2 dB ($p < 0.005$) in the Voiceless geminate. With respect to the release phase (Rel), and looking at the onset, there is an overall statistically significant increase by an average of 4.4 dB ($p < 0.0001$) in the Voiced context. Within geminates vs singletons, there was a statistically significant decrease by 2.4 dB in the Voiced geminate ($p < 0.0001$) and a tendency for the Voiceless geminate to have a decrease in the intensity by 0.6 dB ($p = 0.078$). At the offset of the release (Rel), there is an overall statistically significant increase in the intensity by an average of 1.7 dB ($p < 0.0001$) in the Voiced context. Within geminates vs singletons, there were no statistically significant differences. And finally, with respect to the Intensity at V2 Onset, there is an overall statistically significant increase by an average of 1.1 dB in the intensity (1.3 dB in singleton, $p < 0.005$; and 1 dB in the geminate, $p < 0.01$) in the Voiced context. Within geminates vs singletons, there were no statistically significant differences.

Overall, intensity levels are modest and more restricted to voicing, with Voiced context being associated with higher intensity. Most of the amplitude differences observed here are close to or beyond the JND in amplitude discrimination (Stevens, 1998, pp. 225–226).

3.1.4. Fundamental frequency (Hz)

With respect to f0 (Hz), the statistical and graphical results are presented in Fig. 6 and Table 5 (also see Appendix A).

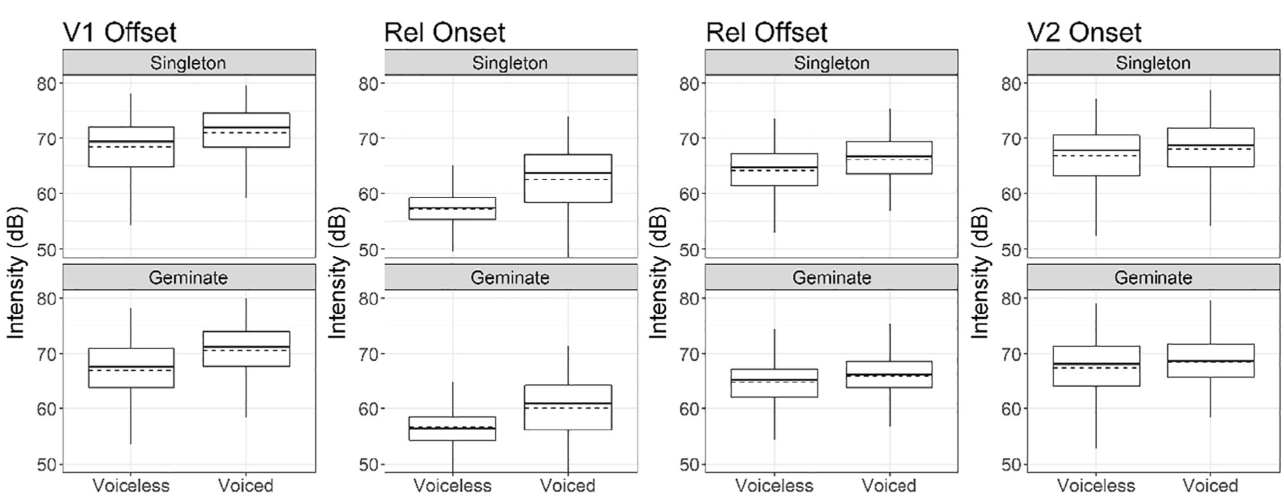

Fig. 5. Intensity (dB) boxplots adjusted by our statistical model for the preceding vowel (V1), the closure phase of the medial consonant (CD) and the following vowel (V2) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

Table 4
Descriptive statistics for the intensity (dB) with mean and standard deviation (SD) for the four-way contrast at the offset of the preceding vowel (V1 Off), at the onset and offset of the release phase (Rel Ons and Rel Off respectively) and at the Onset of the following vowel (V2 Ons).

	Voiceless Singleton		Voiced Singleton		Voiceless Geminate		Voiced Geminate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V1 Off	68.37	5.22	71.12	5.16	67.11	5.27	70.60	5.02
Rel Ons	57.23	3.46	62.56	5.82	56.60	3.49	60.12	5.44
Rel Off	64.08	4.68	66.09	4.93	64.67	4.75	65.99	5.71
V2 Ons	66.75	5.27	68.01	5.42	67.46	5.37	68.43	5.19

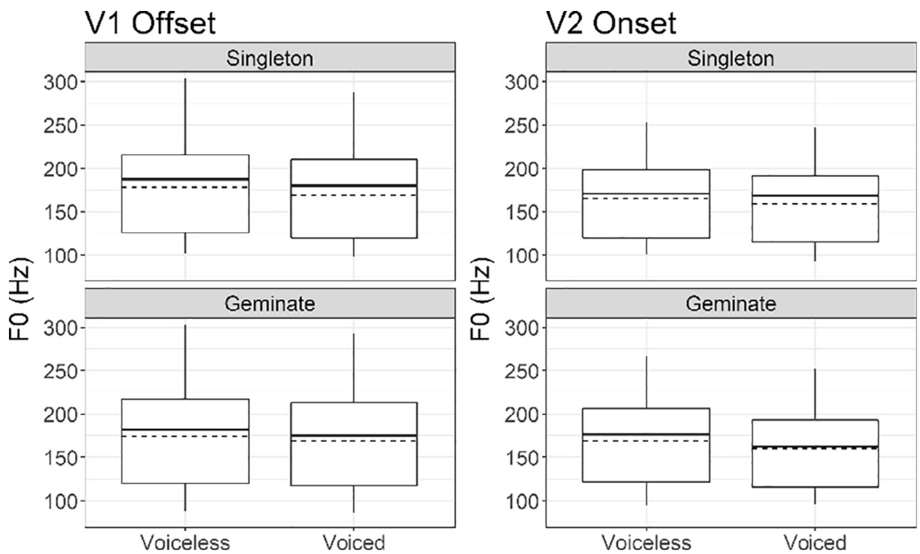

Fig. 6. Fundamental frequency (f0, Hz) boxplots adjusted by our statistical model at the offset of the preceding vowel (V1), and at the onset of the following vowel (V2) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

Table 5
Descriptive statistics for the fundamental frequency (Hz) with mean and standard deviation (SD) for the four-way contrast at the offset of the preceding vowel (V1 Off) and at the Onset of the following vowel (V2 Ons).

	Voiceless Singleton		Voiced Singleton		Voiceless Geminate		Voiced Geminate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V1 Off	177.92	54.38	168.92	51.69	174.33	54.67	169.18	52.76
V2 Ons	165.45	45.46	158.50	43.39	168.81	48.86	160.08	45.86

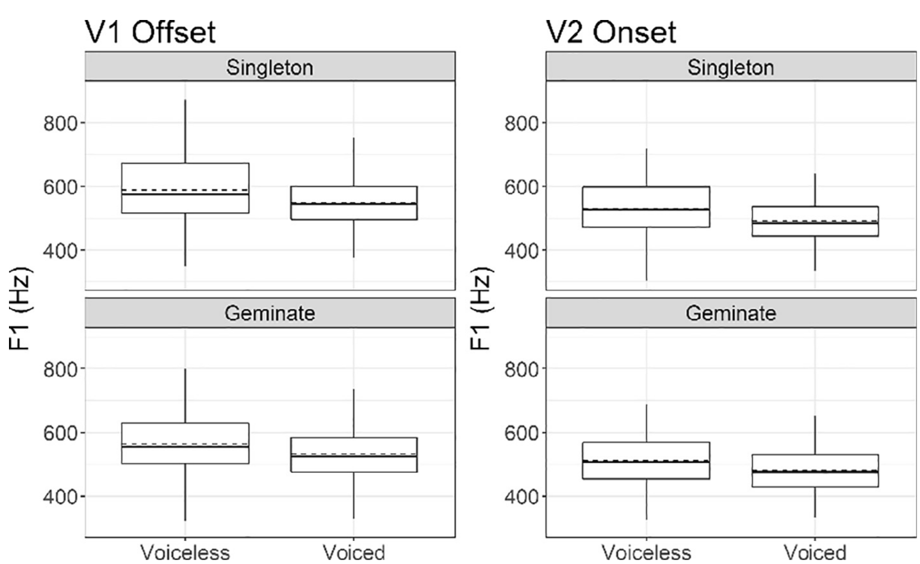

Fig. 7. Frequency of F1 (Hz) boxplots adjusted by our statistical model at the offset of the preceding vowel (V1), and at the onset of the following vowel (V2) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

Starting with V1 Offset, there is a tendency towards a statistically significant decrease by 7 Hz in f0 at V1 Offset ($p = 0.06$) in the Voiced context. Within geminates vs singletons, there were no statistically significant differences. Moving on to V2 Onset, there was a tendency to observe a decrease in the Voiced context by in f0 at V2 Onset by 7 Hz in singleton ($p = 0.07$); and a statistically significant decrease by 9 Hz in the geminate ($p < 0.01$). And there were no statistically significant differences within the singleton vs geminates.

Fundamental frequency differences observed here are close to the JND in pitch discrimination, which is close to 1 Hz difference for complex tones with frequencies between 80 and 500 Hz (Kollmeier, Brand, & Meyer, 2008, p. 65; Stevens, 1998, pp. 227–228).[10]

[10] Although we are aware of different thresholds for pitch discrimination, with Klatt (1973) referring to 2 Hz, and 't Hart (1981) to 1.5 to 2 semitones (9–12 Hz re 100Hz) for pitch movements and up to 3 semitones (1 Hz re 100 Hz) for piano tones.

Table 6
Descriptive statistics for the first formant frequency (Hz) with mean and standard deviation (SD) for the four-way contrast at the offset of the preceding vowel (V1 Off) and at the Onset of the following vowel (V2 Ons).

	Voiceless Singleton		Voiced Singleton		Voiceless Geminate		Voiced Geminate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V1 Off	588.07	106.44	548.74	74.32	564.31	103.32	531.34	79.05
V2 Ons	530.05	93.72	492.36	65.48	511.16	82.37	480.71	67.85

3.1.5. F1 (Hz)

Moving on to F1, the statistical and graphical results are summarised in Table 6 and graphically presented in Fig. 7 (also see Appendix A). Starting with V1 Offset, there is a statistically significant decrease by an average of 36 Hz in F1 at V1 Offset ($p < 0.0001$) in the Voiced context. Within geminates vs singletons, there was a statistically significant decrease in F1 at V1 Offset in the geminate context in the Voiced context by 17.4 Hz ($p < 0.005$) and by 24 Hz in the Voiceless ($p < 0.001$) Moving on to V2 Onset, there is an overall statistically significant decrease by an average of 34 Hz in F1 at V2 Onset ($p < 0.0001$) in the Voiced context.

Within geminates vs singletons, there was a statistically significant decrease in F1 at V2 Onset in the geminate context in the Voiced context by 12 Hz ($p < 0.05$) and in the Voiceless context by 19 Hz ($p < 0.005$). F1 frequency differences observed here are close to the JND in frequency discrimination, which is close to 3 Hz (for frequencies below 500 Hz) and 0.6% for frequencies above 1000 Hz (Kollmeier, et al., 2008, p. 65).

The patterns found for F1 for voicing are comparable with f0, reflecting what is typically reported for a voiceless/fortis-type production (Castleman & Diehl, 1996; Jessen, 2001; Kingston & Diehl, 1995). This also applies to the F1 results in the geminate context, even though no effect was found on f0.

3.1.6. $H1^*–H2^*$ (dB)

Finally, the statistical and graphical results of $H1^*–H2^*$ (dB) are summarised in Table 7 and Fig. 8 (also see Appendix A). At the offset of V1, there are no statistically significant results, either in the model comparisons, or in the pairwise comparisons. At the onset of V2 however the Voiced context showed a tendency for a decrease in $H1^*–H2^*$ by 0.6 dB ($p = 0.054$) for singletons and a marginal statistically significant decrease by 0.58 dB ($p = 0.0497$) for geminates. Within singleton vs geminates, there was a tendency for the geminate to show a decrease in $H1^*–H2^*$ by 0.54 dB ($p = 0.07$) in the Voiceless context, and a marginal statistically significant decrease in $H1^*–H2^*$ by -0.55 dB ($p = 0.0497$) in the Voiced context.

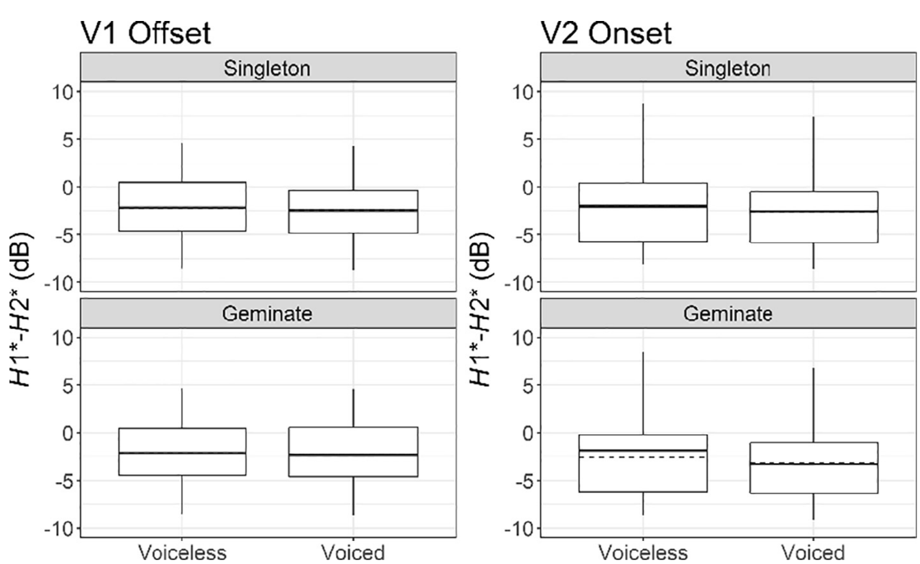

Fig. 8. Harmonic differences $H1^*–H2^*$ (dB) boxplots adjusted by our statistical model at the offset of the preceding vowel (V1), and at the onset of the following vowel (V2) in the Voiceless vs Voiced contexts in the singleton (top) vs geminate (bottom) contexts. The dashed horizontal line in each box represents the mean.

Table 7
Descriptive statistics for $H1^*–H2^*$ (dB) with mean and standard deviation (SD) for the four-way contrast at the offset of the preceding vowel (V1 Off) and at the Onset of the following vowel (V2 Ons).

	Voiceless Singleton		Voiced Singleton		Voiceless Geminate		Voiced Geminate	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V1 Off	−2.22	3.42	−2.42	3.44	−2.13	3.28	−2.27	3.42
V2 Ons	−2.00	4.06	−2.60	3.77	−2.56	4.06	−3.14	3.77

Before summarising the overall results, our aim in the next section is to evaluate the robustness of all acoustic correlates when combined in terms of distinguishing between Voiced and Voiceless stops as well as the singleton and geminate contrast. We report on the results of the exploratory Random Forests analysis below.

3.2. Random forest results

As highlighted in Section 2.4.2, we grew three Random Forests via Conditional Inference Trees. The three models were run with the following specifications: the four-way contrast of voicing by gemination using the full 19 predictors (Model A), 18 predictors without the closure duration (Model B) or 17 without the closure phase or the VOT (Model C). The aim of Model B was to evaluate the impact of taking out the closure duration, a correlate that is usually reported as important for both voicing and gemination, while the aim of Model C was to evaluate the role of the predictors without the closure phase or the VOT.

Starting with the coefficient of correlation, R^2, the results suggest that when the duration of the closure phase was included (Model A), the forest correlated extremely well with the current data as it explained most of the variance in the data ($R^2 = 0.964$). When the duration of the closure phase was not included, (Model B) the correlation fell to $R^2 = 0.802$, and VOT became the main contributor (alongside the other predictors) to a good model fit. Finally, when both the closure duration and the VOT were taken out, the correlation fell to $R^2 = 0.561$. This is an indication that all additional correlates are secondary, as their contribution alone does not highly correlate with the structure of our data.

Classification rates showed the same patterns, with extremely high classification rates of 92.5% for model A; this rate dropped to 82.3% when the closure phase was removed (Model B), which again dropped to 67.2% when both closure duration and VOT were removed (Model C). This again is a clear indication that closure duration is primary, as its inclusion increased the predictive accuracy of our model by 10.2% (comparing Models A and B). The confusion matrices presented in Table 8 show that, for model A, most of confusions were within the Voiceless vs Voiced contexts within each of singleton or geminate categories. The confusions in Model B were variable: when the original data was in a Voiced category, most of the confusions were with the Voiceless (see e.g., Model B columns Vd-S or Vd-G), but when the original data was in a Voiceless category, the confusions were within the singleton or geminates. It is interesting to see that the Voiceless singletons (Model B column Vl-S) were highly confused with the Voiceless geminate (Model B column Vl-G). This result mirrors our finding for the positive VOT's inability to distinguish these two categories (see Section 3.1.2). Moving to Model C, the majority of the confusions were within the Voiced or Voiceless contrasts (e.g., for Voiced singleton, there were confusions with the Voiced geminate at a rate of 26.4%). Comparing Models B and C, it is clear that there are more confusions within the Voiced categories in the latter; the results with the Voiceless categories are comparable.

We then evaluated the contribution of each predictor in the three random forests via the Variable Importance, which was

Table 8
Confusion matrices in percentages of classification results from the three Random Forests (Models A, B and C) with the prediction in rows and original data in columns. Vl = Voiceless, d = Voiced, S = Singleton, G = Geminate. See text for details on models.

	Model A				Model B				Model C			
	Vl-S	Vd-S	Vl-G	Vd-G	Vl-S	Vd-S	Vl-G	Vd-G	Vl-S	Vd-S	Vl-G	Vd-G
Vl-S	**89.5**	7.4	1.5	0	**57.1**	6.8	15	3.3	**51.4**	4.7	16.5	4.7
Vd-S	8.6	**92.6**	0	0	8.6	**91.9**	0	1.4	6.7	**67.6**	0	23.1
Vl-G	0	0	**92.5**	6.1	32.4	0.7	**78.9**	5.2	32.5	1.4	**78.2**	4.2
Vd-G	1.9	0	6	**93.9**	1.9	0.7	6	**90.1**	9.5	26.4	5.3	**67.9**
Total	100	100	100	100	100	100	100	100	100	100	100	100

Bold-faced values are percent correct classification of original data and predicted data.

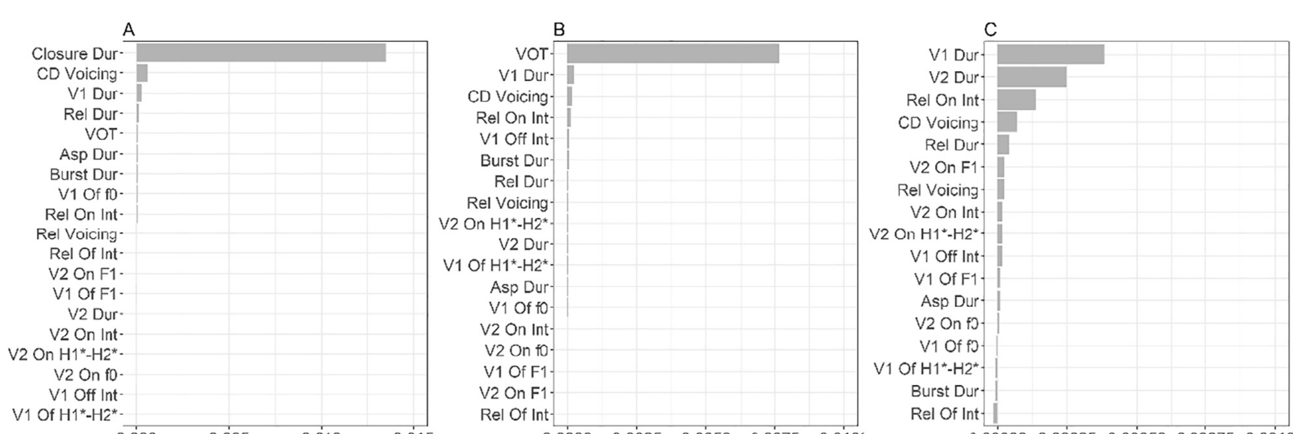

Fig. 9. Mean decrease in accuracy importance scores for the four-way contrast using the full 19 predictors (A), 18 predictors without the closure duration (B) and 17 predictors without the closure duration or the VOT (C). V1 = preceding vowel; V2 = following vowel; CD = closure duration; Rel = Release phase; Asp = Aspiration; On = Onset; Of = Offset; Dur = duration; Int = intensity.

estimated after controlling for correlations and interactions, by using the `varimp` function and `conditional = TRUE`. The graphical results are presented in Fig. 9 and confirm the patterns described above with respect to the closure phase being the most important predictor in model A. This was followed by the proportion of voicing in the closure, preceding vowel duration, release phase duration and then VOT. All remaining predictors partially contributed to the difference, with negative/null scores indicating non-significant ones. With respect to Model B (Fig. 9), VOT took over to become the primary correlate *without* the inclusion of the closure duration, which was then followed by preceding vowel duration, voicing in the closure, then intensity at the onset of the Release phase. It is interesting to see that the VOT in Model A was fifth, indicating that most of the information contained within it is already in the closure and the voicing in the closure. Moving on to Model C, the results show that preceding and following vowel duration, intensity at the onset of the Release, voicing in the closure and duration of the Release phase are the most important predictors; all other predictors contribute to the contrast with variable scores except for the last four. The nature of these predictors is a complex one, with the top two for instance aligning with attributes found in fortis characteristics, while the next two are more characteristic of lenis attributes.

4. Summary and discussion

This paper examined the role of multiplicity of acoustic correlates in evaluating the voicing contrast of stops in the singleton vs geminate contrast in Lebanese Arabic. One main aim was to report on cues for voicing in LA which do not only tap into timing patterns in the consonant (including VOT) and surrounding vowels, but also non-temporal features that can help explore the nature of fortis and lenis characteristics in a voicing language like LA. Another aim was to explore how acoustic correlates of gemination interact with those of voicing and to evaluate the degree to which geminate stops behave in the same manner as fortis/tense stops and how their voicing patterns may interact with their tense features. The combination of the two aims was used to inform whether voicing characteristics can be adequately represented by the [±voice] feature in LA or whether the presence of Voiced geminates requires multiple features and a more graded approach.

A total of 19 acoustic correlates were looked at, and the results are summarised with respect to effects of voicing and gemination. In terms of the voicing contrast, our temporal results place LA firmly into the category of true voice languages, with Voiced stops being prevoiced while Voiceless ones are mainly voiceless unaspirated, as demonstrated by our VOT results (Fig. 4, VOT). This applies to both singleton and geminate stops. Importantly though, our results show that closure duration is the most important correlate for distinguishing the 4-way contrast created by voicing and gemination. While for the Voiced category, closure may overlap with negative VOT, within the Voiceless category, release alone does not distinguish Voiceless singleton and geminate stops. Within the Voiced category, while both singleton and geminate stops pass the 50% voicing criterion in the closure duration to qualify as pre-voiced, actual % voicing can further distinguish between them and shows reduced prevoicing for geminates. This, together with the durational properties of V1 and the release, allows for a clear distinction of the four categories of stops (Voiceless singleton, Voiced singleton, Voiceless geminate, and Voiced geminate).

Looking at medial position further allowed us to explore patterns of active and passive voicing and to contribute to the discussion on language-specific features in the implementation of timing, phonation, and articulatory strength. While in utterance-initial position the voicing contrast is typically carried by the release phase, in medial position the closing movement of the stop is more perceptually salient, leading to the main cues for fortis/lenis to influence the preceding vowel (Castleman & Diehl, 1996; Chen, 1970; Elert, 1964; House & Fairbanks, 1953; Kohler, 1984; Port, 1981; Port, Al-Ani, & Maeda, 1980; Westbury and Keating, 1986). Indeed, the shortening of V1 before geminate stops, together with the presence of passive devoicing in both Voiced singleton and geminate stops, suggests a weaker [voice] feature than say, Russian, which has been shown not to display passive devoicing for Voiced stops (Ringen & Kulikov, 2012). This encourages a graded approach to the implementation of the [voice] feature in medial position, as suggested by Beckman et al. (2013). A numerical value to the [voice] feature in LA might therefore be [8voice] for singletons and [6voice] for geminates. And while Voiceless stops in voicing languages have been found to never exhibit passive voicing (Jansen, 2004), LA Voiceless stops did show a moderate degree of voicing in the closure duration, supporting Beckman et al.'s (2013) claim that laryngeal features are privative rather than binary, i.e. that Voiceless stops in LA show patterns that are more consistent with no laryngeal specification or [Ø] than [spread glottis]. In order to separate the degree of passive voicing in singleton and geminate Voiceless stops in LA, Beckman et al.'s suggestion that more than one laryngeal feature is again helpful, with singleton Voiceless stops being assigned [Ø] [3voice] and geminate Voiceless [Ø] [1voice].

Voiceless geminates exhibited added V1 shortening compared with Voiceless singletons, and Voiced geminates exhibited added devoicing compared with Voiced singletons. Neither contrast, however was neutralised: V1 in Voiced geminates was still longer than in Voiceless geminates, and Voiced geminates managed to retain a good proportion of their voicing. It is tempting to see Voiceless (fortis) stops in LA, especially geminates as having an added minor spread glottis feature; however, the timing patterns in the release did not show aspiration for any of the four categories. Instead, the timing patterns discussed here, along with the non-temporal patterns below, point to a secondary [tense] quality that is voice quality (e.g. creak in geminates) and supraglottal events related to the strength and duration of closure, and its repercussion of other events, e.g. increase in burst duration and longer V2. This would lead to the following possible categorisation: Voiceless singletons: [3voice] [Ø] [2tense]; Voiced singletons: [8voice] [Ø] [0tense]; Voiceless geminates: [1voice] [Øsg] [4tense]; Voiced geminates: [6voice] [Øsg] [3tense]. The numbers are of course relative and their absolute value arbitrary, but an examination of voicing and gemination in all possible word positions would help refine them.

In addition to the shortening and devoicing effects, LA stops exhibited non-temporal patterns which are consistent with the [tense] (or [fortis]) features referred to above (Jessen, 2001; Kohler, 1984). These included higher f0 and F1 in vowels surrounding Voiceless than Voiced contexts, and lower intensity in Voiceless than in Voiced stops in all phases examined (closure, release, V1 offset, V2 offset), with an added effect of geminates for F1. Geminates further exhibited lower $H1^*$–$H2^*$ at the onset of V2, indicating tense or creaky voice quality (Al-Tamimi & Khattab, 2015; Arvaniti & Tserdanelis, 2000; Esposito and DiBenedetto, 1999; Idemaru & Guion, 2008; Lahiri & Hankamer, 1988; Payne, 2005, 2006; Ridouane, 2010; Tserdanelis & Arvaniti, 2001).

The results of the multiple acoustic correlates were then assessed via Random Forests using Conditional Inference Trees and showed a clear advantage for temporal correlates in explaining the contrast. Model A had the highest predictive accuracy at 92.5%, and the duration of the closure phase was always *the* main acoustic correlate used by the Random Forests, followed by voicing in the closure, preceding vowel duration, release duration and then VOT. When the duration of the closure phase was removed (Model B), the VOT became the first correlate to be used by the Random Forests, albeit with a decrease in predictive accuracy (at 82.3%), and more confusions within the Voiceless singleton and geminate stops. Finally, when both the duration of the closure and the VOT were taken out, this allowed the secondary acoustic correlates to be used on their own and they provided a relatively mid-to-high classification rate (at 67.2%). On top of the secondary correlate was preceding and following vowel duration, intensity at onset of the Release and the voicing in the closure.

Our results show that while LA is firmly a true voice language, the [fortis] and [lenis] features that are apparent in each of the voicing and the singleton-geminate contrast interact and require the use of more than one feature to represent the voicing and phonological length contrasts, in the way we have presented above. Durational patterns are clearly very important for both voicing and gemination, something that should not come as a surprise for a language which contrasts phonemic vowel and consonant length. An examination of active and passive voicing patterns, however, along with non-temporal features that are typically associated with a tense articulation, enable us to better present a taxonomy of the 4-way contrast investigated here. While the Voiceless geminates displayed clear fortis properties in terms of their active devoicing and the f0, F1 and $H1$-$H2$ properties of their surrounding vowels, their release patterns remained relatively modest compared with that is typically found in [fortis] stops. Similarly, while Voiced geminates exhibited passive devoicing beyond what was found in their Voiceless counterparts, they retained a high enough proportion to allow for voicing lead and displayed a mixture of attenuated lenis and fortis properties.

In his treatment of the fortis/lenis features, Kohler (1984) saw the fortis/lenis contrast as realised through both articulatory timing and laryngeal power/tension. The first is implemented through speed of stricture formation and release and is considered universal. This study, however, has shown that where there is also a phonological length contrast, speed of constriction formation and release cannot be universal, but are rather determined by the phonology of the language. The second is implemented through aspiration, voicing, and glottalisation. This is also language specific, so that [fortis] can incorporate many other parameters such as [long], [voiced], [aspirated], [tense/stiff], etc. This requires more features to be evoked for phonological systems that distinguish more than two series, and they do not have to be binary.

Acknowledgements

The authors would like to thank the editors of this special issue for choosing our paper, the anonymous reviewers who made comments on a previous version of this paper. This research was supported by an ESRC First Grant (RES-061-23-0053) and a British Academy Overseas Conference Grant 2011.

Appendix A

Full statistical results from the LMM for each of the acoustic correlates with β coefficients (Standard Error, SE), t value (p value; ***=0.0001, **=0.001, *=0.01, =0.05, ns = non-significant). INT = Intercept, Vl = Voiceless, Vd = Voiced, Sing = Singleton, Gem = Geminate, Syll-Type = Syllable type, place = place of articulation, V1 = preceding vowel, CD = closure phase of medial consonant, V2 = following vowel, Rel = Release phase, B = Burst phase, Asp = Aspiration phase, VOT = VOT measure, Ons = Onset, Off = Offset. The table is to be read as follows: a positive β is associated with an increase in the outcome associated with a fixed effect, and vice versa.

Acoustic correlate			Int	Vl-Vd	Sing-Gem	Sex	V1-Length	Syll-Type	Place
Duration	V1	β	102.82	10.57	−10.36	4.04	81.47	−14.92	−2.03
		(SE)	(3.89)	(2.64)	(2.55)	(6.83)	(4.06)	(3.31)	(3.43)
		t	26.41	3.99	−4.07	0.59	20.08	−4.52	−0.59
		(p)	(***)	(***)	(***)	(ns)	(***)	(***)	(ns)
	CD	β	121.25	−22.57	101.66	9.82	1.66	−7.41	−10.94
		(SE)	(3.1)	(2.25)	(4.21)	(5.67)	(2.01)	(2.97)	(2.66)
		t	39.06	−10.02	24.16	1.73	0.83	−2.5	−4.11
		(p)	(***)	(***)	(***)	(ns)	(ns)	(*)	(***)
	V2	β	182.94	7.47	15.92	−12	4.22	94.46	−3.28
		(SE)	(6.27)	(3.88)	(3.58)	(11.47)	(3.87)	(6.93)	(4.75)

(*continued on next page*)

(*continued*)

Acoustic correlate			Int	Vl-Vd	Sing-Gem	Sex	V1-Length	Syll-Type	Place
		t	29.16	1.92	4.45	−1.05	1.09	13.63	−0.69
		(*p*)	(***)	(·)	(***)	(ns)	(ns)	(***)	(ns)
	Rel	β	24.97	−9.39	0.34	1.64	5.14	1.55	12.39
		(SE)	(1.25)	(1.33)	(1.01)	(2.11)	(1.08)	(1.21)	(1.67)
		t	19.99	−7.06	0.34	0.78	4.76	1.28	7.44
		(*p*)	(***)	(***)	(ns)	(ns)	(***)	(ns)	(***)
	B	β	8.85	−2.12	0.99	1.05	0.34	0.45	4.15
		(SE)	(0.35)	(0.33)	(0.23)	(0.63)	(0.26)	(0.3)	(0.6)
		t	25.37	−6.46	4.26	1.68	1.33	1.49	6.94
		(*p*)	(***)	(***)	(***)	(ns)	(ns)	(ns)	(***)
	Asp	β	16.25	−7.17	−1.08	0.25	4.57	1.06	8.04
		(SE)	(1.16)	(1.21)	(0.89)	(1.98)	(0.98)	(1.07)	(1.53)
		t	14.05	−5.91	−1.21	0.12	4.65	0.99	5.26
		(*p*)	(***)	(***)	(ns)	(ns)	(***)	(ns)	(***)
Voicing	VOT	β	−42.31	−115.23	−45.27	6.46	10.9	5.19	6.7
		(SE)	(6.13)	(8.36)	(7.05)	(8.52)	(6.3)	(7.44)	(8.38)
		t	−6.91	−13.78	−6.43	0.76	1.73	0.7	0.8
		(*p*)	(***)	(***)	(***)	(ns)	(·)	(ns)	(ns)
	CD	β	58.37	63.21	−14.47	−9.77	−5.58	1.69	5.56
		(SE)	(2.26)	(3.63)	(1.61)	(4.23)	(1.74)	(1.8)	(2.02)
		t	25.82	17.43	−8.99	−2.31	−3.21	0.94	2.75
		(*p*)	(***)	(***)	(***)	(*)	(**)	(ns)	(**)
	Rel	β	38.64	63.17	−23.56	−20.69	−7.87	8.7	0.76
		(SE)	(2.57)	(5.78)	(3.28)	(3.99)	(2.93)	(3.24)	(3.22)
		t	15.03	10.94	−7.19	−5.18	−2.69	2.7	0.24
		(*p*)	(***)	(***)	(***)	(***)	(*)	(**)	(ns)
Intensity	V1Off	β	67.35	2.36	0.16	−5.64	−1.52	−3.78	−1.86
		(SE)	(0.94)	(0.35)	(0.26)	(1.84)	(0.33)	(0.42)	(0.38)
		t	71.59	6.67	0.62	−3.06	−4.59	−9.06	−4.87
		(*p*)	(***)	(***)	(ns)	(**)	(***)	(***)	(***)
	RelOns	β	59.33	5.44	−2.35	−3.8	−2.6	0.09	1.9
		(SE)	(0.84)	(0.81)	(0.43)	(1.6)	(0.42)	(0.56)	(0.67)
		t	70.29	6.8	−5.43	−2.37	−6.13	0.17	2.84
		(*p*)	(***)	(***)	(***)	(*)	(***)	(ns)	(**)
	RelOff	β	64.6	1.28	−0.74	−4.36	−2.45	0.81	−1.65
		(SE)	(0.92)	(0.34)	(0.25)	(1.81)	(0.27)	(0.39)	(0.36)
		t	70.27	3.83	−2.93	−2.41	−9.09	2.11	−4.58
		(*p*)	(***)	(***)	(**)	(*)	(***)	(*)	(***)
	V2Ons	β	67.19	0.79	−0.45	−6.06	−2.49	0.99	−1.48
		(SE)	(0.95)	(0.38)	(0.26)	(1.87)	(0.25)	(0.4)	(0.36)
		t	70.68	2.08	−1.69	−3.23	−10.09	2.47	−4.09
		(*p*)	(***)	(*)	(·)	(**)	(***)	(*)	(***)
F0	V1Off	β	161.42	−4.66	4.56	89.85	−12.21	−21.75	0.4
		(SE)	(4.74)	(1.41)	(1.16)	(9.36)	(2.53)	(2.34)	(1.48)
		t	34.06	−3.31	3.95	9.6	−4.83	−9.31	0.27
		(*p*)	(***)	(**)	(***)	(***)	(***)	(***)	(ns)
	V2Ons	β	166.59	−4	−3.42	86.82	−3.55	15.31	3.12
		(SE)	(4.39)	(1.81)	(1.54)	(8.61)	(1.75)	(3.6)	(1.98)
		t	37.97	−2.21	−2.22	10.09	−2.03	4.25	1.58
		(*p*)	(***)	(*)	(*)	(***)	(*)	(***)	(ns)
F1	V1Off	β	500.36	−57.29	−1.32	70.56	−52.44	−85.4	−64.58
		(SE)	(13.11)	(15.01)	(11.69)	(19.13)	(13.66)	(19.56)	(17.67)
		t	38.16	−3.82	−0.11	3.69	−3.84	−4.37	−3.66
		(*p*)	(***)	(***)	(ns)	(**)	(***)	(***)	(***)
	V2Ons	β	451.22	−48.76	−13.84	66.26	−82.62	−55.58	−61.51
		(SE)	(10.55)	(10.74)	(8.39)	(16.77)	(9.68)	(11.72)	(13.79)
		t	42.78	−4.54	−1.65	3.95	−8.54	−4.74	−4.46
		(*p*)	(***)	(***)	(ns)	(***)	(***)	(***)	(***)

(*continued*)

Acoustic correlate			Int	Vl-Vd	Sing-Gem	Sex	V1-Length	Syll-Type	Place
H1–H2**	V1Off	β	−2.04	0.06	0	1.83	0.08	0.34	0.4
		(SE)	(0.8)	(0.28)	(0.25)	(1.56)	(0.26)	(0.32)	(0.36)
		t	−2.55	0.19	−0.01	1.17	0.32	1.06	1.13
		(*p*)	(*)	(ns)	(ns)	(ns)	(ns)	(ns)	(ns)
	V2Ons	β	−2.29	−0.19	−0.72	2.69	−0.25	0.34	0.81
		(SE)	(0.9)	(0.34)	(0.29)	(1.76)	(0.28)	(0.35)	(0.46)
		t	−2.56	−0.56	−2.46	1.53	−0.92	0.96	1.74
		(*p*)	(*)	(ns)	(*)	(ns)	(ns)	(ns)	(·)

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2018.09.010.

References

Abramson, A. S. (1986). The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association, 16*(1), 8–16. https://doi.org/10.1017/S0025100300003054.

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002.

Al-Ani, S. (1970). *Arabic phonology: An acoustical and physiological investigation*. Paris: The Hague Mouton. https://doi.org/10.1515/9783110878769.

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 8*(1), 1–40. https://doi.org/10.5334/labphon.19.

Arvaniti, A. (1999). Effects of speaking rate on the timing of single and geminate sonorants. In Proceedings of the XIVth international congress of phonetic sciences (pp. 599–602). San Francisco, CA.

Al-Tamimi, J., & Khattab, G. (2015). Acoustic cue weighting in singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants. *Journal of the Acoustical Society of America, 138*(1), 344–360. https://doi.org/10.1121/1.4922514.

Al-Tamimi, J. (2004). L'équation du locus comme mesure de la coarticulation VC et CV : Étude préliminaire en Arabe Dialectal Jordanien. In Actes des XXVèmes Journées d'Étude sur la Parole (Vol. 9, pp. 7–10). Retrieved from http://aune.lpl.univ-aix.fr/jep-taln04/proceed/actes/jep.htm.

Al-Tamimi, J. (2007). Static and Dynamic cues in vowel production: a cross dialectal study in Jordanian and Moroccan Arabic. In Proceedings of the 16th international congress of phonetic sciences (ICPhS), Saarbrücken, Germany (pp. 541–544).

Aoyama, K. (2002). Quantity contrasts in Japanese and Finnish: Differences in adult production and acquisition. In Studies in language sciences (2): Papers from the second annual conference of the japanese society for language sciences (pp. 121–135).

Arvaniti, A. (2001). Comparing the phonetics of single and geminate consonants in Cypriot and Standard Greek. In *Proceedings of the 4th International Conference on Greek Linguistics* (pp. 37–44). Thessaloniki: University Studio Press.

Arvaniti, A., & Tserdanelis, G. (2000). On the phonetics of geminates: evidence from Cypriot Greek. In Proceedings of 6th international conference on spoken language processing (Vol. 2, pp. 559–562). Beijing, China.

Arvaniti, A. (1999). Effects of speaking rate on the timing of single and geminate sonorants. In Proceedings of the XIVth international congress of phonetic sciences (pp. 599–602). San Francisco, CA.

Auguie, B. (2016). gridExtra: Miscellaneous functions for "Grid" graphics. Retrieved from https://cran.r-project.org/package=gridExtra.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Palgrave Macmillan.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48 https://doi.org/10.18637/jss.v067.i01.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics, 70*, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics, 49*(2), 259–284. https://doi.org/10.1017/S0022226712000424.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boersma, P., & Weenink, D. (2009). Praat. Doing Phonetics by Computer. V. 5.1.0: Retrieved 22 July 2009 from http://www.praat.org/.

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics, 66*, 45–60. https://doi.org/10.1016/j.pragma.2014.02.011.

Butcher, A. (2004). Fortis/lenis'' revisited one more time: The aerodynamics of some oral stop contrasts in three continents. *Clinical Linguistics & Phonetics, 18*(6–8), 547–557. https://doi.org/10.1080/02699200410001703565.

Castleman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics, 24*(4), 383–398. https://doi.org/10.1006/jpho.1996.0021.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22*(3), 129–159. https://doi.org/10.1159/000259312.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cohn, A., Ham, W., & Podesva, R. (1999). The phonetic realization of singleton-geminate contrasts in three languages of Indonesia. In Proceedings of the 14th ICPhS (pp. 587–590).

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 837–845*. https://doi.org/10.2307/2531595.

Deterding, D., & Nolan, F. (2007). Aspiration and voicing of Chinese and English plosives. In Proceedings of the 16th international congress of phonetic Sciences (pp. 385–388).

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Walter de Gruyter.

Elert, C. C. (1964). *Phonological studies of quantity in Swedish*. Uppsala: Almqvist & Wiksell.

Esposito, A., & DiBenedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *Journal of the Acoustical Society of America, 106*(4, Pt. 1), 2051–2062. https://doi.org/10.1121/1.428056.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*(2), 125–146.

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics, 40*(1), 152–161. https://doi.org/10.1016/j.wocn.2011.10.003.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT Quarterly Progress Report, 101*, 198–212.

Ham, W. (2001). *Phonetic and phonological aspects of geminate timing*. Routledge.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics, 29*, 451–480 https://doi.org/10.006/jpho.2001.0146.

Harris, J. (1994). *English sound structure*. Blackwell Oxford.

Hassan, Z. M. (2002). Gemination in Swedish and Arabic with a particular reference to the preceding vowel: an instrumental and comparative approach. In Proceedings of Fonetik TMH-QPSR (Vol. 44, pp. 81–85).

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Berlin: Springer Series in Statistics Springer.

Hawks, J. W., & Miller, J. D. (1995). A formant bandwidth estimation procedure for vowel synthesis. *Journal of the Acoustical Society of America, 97*(2), 1343–1344. https://doi.org/10.1121/1.412986.

Hayes, B., & Steriade, D. (2004). Introduction: The phonetic bases of phonological markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 1–33). Cambridge: Cambridge University Press.

Hirst, D. (2011). The analysis by synthesis of speech melody: From data to models. *Journal of Speech Sciences, 1*, 55–83.

Homma, Y. (1981). Durational relations between Japanese stops and vowels. *Journal of Phonetics, 9*, 273–281.

Honeybone, P. (2005). Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. In M. van Oostendorp & J. van de Weijer (Eds.), *The internal organization of phonological segments* (pp. 319–354). Berlin: Mouton de Gruyter.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics, 15*(3), 651–674. https://doi.org/10.1198/106186006X133933.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*(1), 105–113. https://doi.org/10.1121/1.1906982.

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association, 38*, 167–186. https://doi.org/10.1017/S0025100308003459.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America, 121*(4), 2283–2295. https://doi.org/10.1121/1.2697522.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology, 12*(3), 369–396. https://doi.org/10.1017/S0952675700002566.

Jaeger, J. (1983). The fortis/lenis question: Evidence from Zapotec and Jawoñ. *Journal of Phonetics, 11*, 177–189.

Jaeger, J. (1978). Speech aerodynamics and phonological universals. In Annual meeting of the Berkeley Linguistics Society (Vol. 4, pp. 312–329).

James, G., Witten, D., & Hastie, T. (2013). An Introduction to Statistical Learning: With Applications in R. *Springer*. https://doi.org/10.1007/978-1-4614-7138-7.

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. University Library Groningen.

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. John Benjamins Publishing.

Jessen, M. (2001). Phonetic Implementation of the Distinctive Auditory Features [voice] and [tense] in Stop Consonants. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 237–294). Mouton de Gruyter.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology, 19*(2), 189–218.

Kawahara, S. (2015). The phonetics of sokuon, obstruent geminates. In H. Kubonozo (Ed.), *The handbook of Japanese language and linguistics: Phonetics and phonology* (pp. 43–73). Mouton.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 286–319*. https://doi.org/10.2307/413642.

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th international congress of phonetic sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0821. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0821.pdf.

Khattab, G., & Al-Tamimi, J. (2008). Durational Cues for Gemination in Lebanese Arabic. *Language and Linguistics, 22*, 39–55.

Khattab, G., & Al-Tamimi, J. (2014). Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure. *Laboratory Phonology*. https://doi.org/10.1515/lp-2014-0009.

Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics. Volume XIII-XIV: Stanford, 1999 and Berkeley, California 2000, Edited by Dilworth B. Parkinson and Elabbas Benmamoun (pp. 1–37).

Kim, M. R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*. https://doi.org/10.1006/jpho.2001.0152.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*, 419–454. https://doi.org/10.2307/416481.

Kingston, J., & Diehl, R. L. (1995). Intermediate properties in the perception of distinctive feature values. *Papers in Laboratory Phonology, 4*, 7–27.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics, 36*(1), 28–54.

Kingston, J., Kawahara, S., Chambless, D., Masha, D., & Brenner-Alsopa, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics, 37*(3), 297–320. https://doi.org/10.1016/j.wocn.2009.03.007.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America, 53*(1), 8–16. https://doi.org/10.1121/1.1913333.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America, 87*, 820–856. https://doi.org/10.1121/1.398894.

Kohler, K. (1984). The phonetic explanation in phonology: The feature fortis/lenis. *Phonetica, 41*, 150–174. https://doi.org/10.1159/000261721.

Kollmeier, B., Brand, T., & Meyer, B. (2008). Perception of speech and sound. In *Springer Handbook of Speech Processing* (pp. 61–82). Springer.

Kuang, J., & Keating, P. (2012). Glottal articulations of phonation contrasts and their acoustic and perceptual consequences Retrieved from. *UCLA Working Papers in Phonetics, 111*, 123–161.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Kulikov, V. (2016). Voicing in Qatari Arabic: Evidence for prevoicing and aspiration. In Qatar Foundation Annual Research Conference Proceedings (Vol. 2016, p. SSHAPP2330).

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Cambridge, MA: Blackwell.

Lahiri, A., & Hankamer, J. (1988). The timing of geminate consonants. *Journal of Phonetics, 16*, 327–338.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830.

Local, J., & Simpson, A. (1999). Phonetic implementation of geminates in Malayalam nouns. In Proceedings of the 14th ICPhS (pp. 595–598).

Local, J., & Simpson, A. (1988). The domain of gemination in Malayalam. In D. Bradley, E. J. A. Henderson, & M. Mazaudon (Eds.), *Prosodic analysis and Asian linguistics: To honour R. Sprigg* (pp. 33–42). Pacific Linguistics Series C-104.

Louali, N., & Puech, G. (1994). Les consonnes 'Fortes' du berbère: Indices perceptuels et corrélats phonétiques. In XXémes JEP (Journées d'Étude sur la Parole), Trégastel, Juin 1994 (pp. 459–464).

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics, 37*(1), 29–45. https://doi.org/10.1016/j.wocn.2008.08.002.

Nasr, R. T. (1960). Phonemic length in Lebanese Arabic. *Phonetica, 5*, 209–211. https://doi.org/10.1159/000258058.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech* (pp. 189–216). Springer.

Ohala, J. J. (1997). Aerodynamics of phonology. In Proceedings of the Seoul International Conference on Linguistics (Vol. 92, p. 97).

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? In International workshop on machine learning and data mining in pattern recognition (pp. 154–168).

Ouakrim, O. (2003). A Perceptual Study of Tenseness. Some Acoustic Cues Identifying Tense Vs Non-Tense Contrast in Berber. In XV internation congress on phonetic sciences (pp. 795–798). Barcelona, Spain.

Payne, E. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association, 35*(2), 153–181. https://doi.org/10.1017/S0025100305002240.

Payne, E. (2006). Non-durational indices in Italian geminate consonants. *Journal of the International Phonetic Association, 36*(1), 83–95. https://doi.org/10.1017/S0025100306002398.

Pellegrino, F., & André-Obrecht, R. (2000). Automatic language identification: An alternative approach to phonetic modelling. *Signal Processing, 80*(7), 1231–1244. https://doi.org/10.1016/S0165-1684(00)00032-3.

Perkell, J. S. (1971). Physiology of speech production: A preliminary study of two suggested revisions of the features specifying vowels. In *Quarterly Progress Report No. 102, Research Laboratory of Electronics* (pp. 123–139). M.I.T.

Podesva, R. (2000). Constraints on geminates in Buginese and Selayarese. In Proceedings of the west coast conference on formal linguistics, Vol. 19 (pp. 343–356).

Port, R. F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America, 69*(1), 262–274. https://doi.org/10.1121/1.385347.

Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica, 37*(4), 235–252. https://doi.org/10.1159/000259994.

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/.

Ridouane, R. (2007). Gemination in Tashlhiyt Berber: An acoustic and articulatory study. *Journal of the International Phonetic Association, 37*(2), 119–142. https://doi.org/10.1017/S0025100307002903.

Ridouane, R. (2010). Geminates at the junction of phonetics and phonology. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Papers in Laboratory Phonology 10* (pp. 61–90). Mouton: de GRUYTER.

Ringen, C., & Kulikov, V. (2012). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics, 20*(2), 269–286. https://doi.org/10.1353/jsl.2012.0012.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1*(2), 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420. https://doi.org/10.1093/beheco/arn145.

Shadle, C. H., Nam, H., & Whalen, D. H. (2016). Comparing measurement errors for formants in synthetic and natural vowels. *Journal of the Acoustical Society of America, 139*(2), 713–727. https://doi.org/10.1121/1.4940665.

Shimizu, K. (1996). *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido.

Solé, M.-I. (2009). Acoustic and aerodynamic factors in the interaction of features. *Phonetics and Phonology: Interactions and Interrelations, 306*, 205. https://doi.org/10.1075/cilt.306.10sol.

Stevens, K. N. (1998). *Acoustic phonetics*. MIT Press.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*(1), 1–11. https://doi.org/10.1186/1471-2105-9-307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 1–25. https://doi.org/10.1186/1471-2105-8-25.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*(4), 323–348. https://doi.org/10.1037/a0016973.

't Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America, 69*(3), 811–821. https://doi.org/10.1121/1.385592.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change, 24*(2), 135–178. https://doi.org/10.1017/S0954394512000129.

Tserdanelis, G., & Arvaniti, A. (2001). The acoustic characteristics of geminate consonants in Cypriot Greek. In Proceedings of the fourth international conference on greek linguistics (pp. 29–36).

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73*(4), 1322–1336. https://doi.org/10.1121/1.389236.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics, 22*(1), 145–166. https://doi.org/10.1017/S0022226700010598.

Wetzels, W. L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language*, 207–244.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. Retrieved from http://ggplot2.org.

Yeni-Komshian, G. H., Caramazza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5*(1), 35–48.